Full Waveform Inversion for Diffraction Energy

A DISSERTATION

submitted towards the partial fulfillment of the

requirements for the award of the degree of

INTEGRATED MASTERS IN TECHNOLOGY

in

GEOPHYSICAL TECHNOLOGY

Submitted by:

HARPREET SINGH SETHI

Enrollment : 11411012

DEPARTMENT OF EARTH SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY, ROORKEE

ROORKEE - 247667, INDIA

May 2015

Candidate's Declaration

I hereby declare that the work which is presented in this dissertation report titled **"Full Waveform Inversion using Diffraction Energy"** in partial fulfilment of the requirements for the award of the degree of **"INTEGRATED MASTER OF TECHNOLOGY"** in **GEOPHYSICAL TECHNOLOGY**, submitted to the Department of Earth Sciences, Indian Institute of Technology Roorkee is an authentic record of my own work carried out during the period from January 2014 to April 2014 under the supervision of Dr. S. Mukopadhyay, Professor, Department of Earth Sciences, Indian Institute of Technology Roorkee, Prof Tariq Alkhalifah, Department of Physical Sciences and Engineering , KAUST , Saudi Arabia.

I have not submitted the matter embodied in this dissertation report for the award of any other degree.

DATE : 28 APRIL 2016 **Harpreet Singh Sethi (11411012)**

Place : IIT Roorkee IMT Geophysical Technology

This is to certify that the above statement made by the candidate is correct to the best of my knowledge and belief . I wish him a great success in life and bright future.

DATE : **Signature of the Guide**

Place : IIT Roorkee Prof. Sagarika Mukhopadhyay

CERTIFICATE

I.....**HARPREET SINGH SETHI**....hereby solemnly declare that the dissertation entitled ...**FULL WAVEFORM INVERSION FOR DIFFRACTION ENERGY**....being submitted by me towards the partial fulfillment of the requirements for the award of ...**INTEGRATED M.TECH GEOPHYSICAL TECHNOLOGY**...degree is a record of my own work and that I have not copied the work of any other person(s) including published literature and material from any web site.

Where ever the work of other(s) has been used, it has been duly acknowledged and quoted with proper reference to the original wok.

I fully understand the implications of plagiarism and that if at any stage the above statement made by me is found to be incorrect; I shall be fully responsible for my act(s)

(Signature .....)

Name.....HARPREET SINGH SETHI.....

Place.....ROORKEE.....

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my thesis supervisors **Prof Sagarika Mukhopadhyay**, Department of Earth Sciences, IIT Roorkee and **Prof Tariq AlKhalifah** ,Physical Sciences and Engineering, KAUST ,SAUDI Arabia.

I am thankful to **Prof Tariq** for inviting me to work with his group . I am highly indebted for his kind gesture to allow me to attend his two courses namely **Seismic Imaging** and **Mathematical Methods in Seismic Imaging** , taking part in group activities including talks, seminars and presentations. He is truly one of the sages of seismic anisotropy and he has imparted me with a huge of amount of knowledge in this short period of time. He always keeps his students inspiring to be aware of the current research scenarios and is expert in bringing best out of them. I have learned a lot from him and there is a lot more to learn but first I want to learn his secret of extraordinary time management skills **Prof Sagarika** has been a constant support of motivation for me to learn something beyond the scope of class notes and lectures .Her constant constructive criticism has helped me to refine my work and develop the skills what I have today.

I am also grateful to **Zedong Wu** and **Yunseok Choi** for the discussions on **Forward Modeling**. I am also thankful to **Ramzi Djebbi**,**Mahesh Kalita**,**Mrinal Sinha**(from Prof Schuster group) and **Gaurav Dutta** (from Prof Schuster group) for their constant help and discussions.

Not to mention , the parents and the omnipresent creator, with their blessings and grace I have gained so much and will continue to gain more.

Abstract

The identification of the small scale geological heterogeneities like pinchouts,faults ,fractures etc is a primary problem in seismics.They hold a significant importance in understanding the geology of a particular area as they can act as structural traps for accumulation of hydrocarbons or can behave as migration pathways aiding the escape of the hydrocarbons.

Standard approaches to obtain this high-resolution information, such as coherency analysis and structure-oriented filters, derive attributes from stacked, migrated images are image-driven.These techniques are sensitive to artifacts due to an inadequate migration velocity; in fact the attribute derivation is even not based on the physics of wave propagation.

These small scale features are usually encoded in the form of diffractions in our seismic data.Thus a seismic section containing only diffractions could be of great value to the interpreter .Diffractions are the other coherent events present along with the reflections in our seismic data.

There are some fundamental differences between reflections and diffractions like different moveout, amplitude etc which can facilitate there separation and here we discuss two methods to focus and separate diffractions. The first part of the thesis is devoted to exploiting the difference between diffractions and reflections on the basis of Snell's law . We try to utilise this to design a weighting function which account this difference in behaviour to directly clean the gradient in Full waveform Inversion.

The second part of the thesis proposes separating the diffractions using a dip-steering filter in the model domain in the Born framework which could be directly inverted using Full waveform inversion.

The algorithms developed in this thesis are coded using GPU's (CUDA C) in a parallel environment except the dip-steering filter algorithm.

Contents

1 Introduction **13**

2 Seismic Diffractions and Imaging **15**

2.0.1 Wave Propagation 15

2.0.2 Reflections and Diffractions 16

2.0.3 Diffraction Imaging and Specularity 20

3 Full Waveform Inversion **22**

3.0.4 Forward Problem 23

3.0.5 Inverse Problem 37

3.0.6 Elements of practical FWI 43

3.0.7 CUDA and Parallelisation strategies in FWI 57

4 Numerical results : FWI **66**

5 Specularity + FWI **70**

5.0.8 Ray based Specularity 72

5.0.9 Wave based Specularity 80

5.1 Discussion 90

6 Diffraction Inversion **92**

6.0.1 Scattering series and Born approximation 92

6.0.2 Diffraction Free model 95

6.0.3 Numerical Results: Diffraction Inversion 96
6.1 Conclusion and Future Work 101

Appendices **102**

A The Adjoint State Method **103**

B Reverse Time Migration **107**

List of Figures

2.1 Velocity model 18

2.2 Finite difference modeling 18

2.3 Reflectivity Model 19

2.4 Zero Offset Modeling 19

2.5 Reflections 20

2.6 Diffractions 20

3.1 Central Difference coefficients for Derivative (source:wikipedia) 25

3.2 Numerical vs Analytical solution 1-D wave equation (source:Fichtner [1]) . 27

3.4 Numerical Dispersion for exact wavespeed $v = 5km/s$ (source:Fichtner [1]) 30

3.5 $sinc[t - 4.0]$ 32

3.6 $Ricker, \nu_0 = 10Hz$ 32

3.7 Reflection coefficients vs Incidence angle (source:Fichtner [1]) 35

3.8 Left:m_1 is global minima and m_2 is local minima Right: m_3 is strict global minima and m_4 is strict local minima (source : Fichtner [1]) 38

3.9 Example of the points satisfying the imaging condition $t_{S,x} + t_{x,R} = T$, for a reflection residual at time T = 5s, in a model with v = 2m/s, with a source at $x_S = 0$ and a receiver at $x_R = 8$ m, resulting in an offset of D = 8m. The points form an ellipse with focal points at the source and receiver positions. 46

3.10 a) $Au = s$, source s generates a wavefield that is propagated by operator A b) $A\frac{\partial u}{\partial v} = -\frac{\partial A}{\partial v}u$, source $-\frac{\partial A}{\partial v}u$ generates diffracted wavefield $\frac{\partial u}{\partial v}$ propagated by operator A 47

3.11 Figure taken from (Operto et al., 2013) [2]. The true velocity model consists of an homogeneous background m_0 and three scattering points. There is one source denoted by a star, and receivers at the positions of the dashed line. b) The source s generates a wavefield that travels through the model m and is going to be diffracted by each of the heterogeneities. a) The partial derivative wavefield $\frac{\partial u}{\partial m}$ (diffracted by the green scatterer) at the receiver positions. c) The data residuals recorded at the receiver positions. Three arrivals are seen due to the three diffraction points. The sum over time of the correlation the partial derivative wavefields (figure a) and the data residuals (figure c) is going to generate a non-zero gradient value at point v_i .d) The role of the acquisition can be analyzed, considering the angle between the source and the receiver. 48

3.12 Figure from Pratt et al. (1998)[3]. With one source and receivers on the surface. On the left, gradient in the first iteration shows the smile artefacts. On the right, the Gauss-Newton descent direction has removed the smile artefacts. 50

3.13 a) Imaging kernel in the frequency domain for one source receive pair. The first central ellipse denotes the first Fresnel zone. The black and white fringes describe the iso-phase surfaces of all the scattered arrivals. (Each isochrone represents all the parameters in the model space that can explain a seismic phase arrival at time t). The width of the isochrones decreases in depth as the scattering angle decreases (Figure from : Woodward, 1992)[4] 52

3.14 Figure adapted from Sirgue and Pratt (2004)[5]. $k_{S,x}$ and $k_{R,x}$ are vectors from the source to the diffracting point and from the receiver to the diffracting point. 54

3.15 Figure from Sirgue (2003)[6]. Under the plane wave approximation, an illustration of the variation of the wave number contribution in the gradient, for different incidence-scattering angles. The first column represents the wave paths from the source to the scattering point $\exp(i\vec{k}.\hat{s})$. The second column represents the wave paths from the receiver to the scattering point $\exp(i\vec{k}.\hat{r})$. The last column is the multiplication of both paths $\exp(i\vec{k}.(\hat{s}+\hat{r})$, which is proportional to the gradient. a) The source are receiver are at the same position. The incidence and scattering angle is 0^0. The total wave number is high, which can be seen in the high resolution of the gradient. b) The incidence and scattering angle is 90^0 . c) The incidence and scattering angle is 160^0 . The lowest wave number contribution is given the high incidence and scattering angles. The high wave number contribution is given by small incidence and scattering angles 55

3.16 Figure from Mora (1989)[7]. a) Shows the part of the usual part of wave number spectrum in the image space that can be resolved with the infinite offset data. As it is clear, there is a gap in low number spectrum. b) Shows the part of the wave number spectrum in the image that can be recovered by introducing reflectivity information in the model, and expanding the imaging condition , in forward and back-scattered direct and back-propagated wavefields. 56

3.17 Imaging condition with and without a reflector in the background model. a) correlation of u and λ. Reconstructs high magnitude wavenumbers. b) Correlation of back-scattered u and λ. Reconstructs low magnitude wavenumbers. 57

3.18 CPU vs GPU design (source:[8]) . . . 59
3.19 CUDA Memory model (source:[8]) . . . 61
3.20 CUDA Memory types (source:[8]) . . . 62
3.21 Red block size shows the shared memory . . . 63
3.22 Wavefield Reconstruction . . . 64
3.23 Parallel Reduction . . . 65

4.1 Marmousi Velocity Model, (IFP) . . . 66
4.2 Initial Velocity Model . . . 67
4.3 Frequencies involved in inversion . . . 67
4.4 Velocity Model recovered after 100^{th} iteration . . . 68
4.5 Velocity Model recovered after 250^{th} iteration . . . 68
4.6 Normalised Objective function value . . . 69

5.5 Velocity model . . . 74
5.6 RTM image . . . 74
5.7 Specularity image . . . 75
5.8 Specularity image . . . 75
5.9 Gradient . . . 76
5.10 Cleaned Gradient after 1^{st} iteration . . . 76
5.11 Velocity model . . . 77
5.12 RTM image . . . 77
5.13 Specularity image . . . 78
5.14 Specularity image . . . 78
5.15 Gradient . . . 79
5.16 Cleaned Gradient after 1^{st} iteration . . . 79
5.17 Velocity model . . . 81
5.18 RTM image . . . 81

5.19 Specularity image 82
5.20 Specularity image 82
5.21 Gradient 83
5.22 Cleaned Gradient after 1^{st} iteration 83
5.23 Velocity model 84
5.24 RTM image 84
5.25 Specularity image 85
5.26 Specularity image 85
5.27 Gradient 86
5.28 Cleaned Gradient after 1^{st} iteration 86
5.29 Velocity model 87
5.30 RTM image 87
5.31 Specularity image 88
5.32 Specularity image 88
5.33 Gradient 89
5.34 Cleaned Gradient after 1^{st} iteration 89
5.35 Velocity Model after 60_{th} iteration 90

6.1 Dip steering radius or width (source:www.opendtect.com) 95
6.2 Point Scatterer model 96
6.3 Background Model 97
6.4 Perturbed Model 97
6.5 Diffractions 98
6.6 Mamrousi model with scatterers 98
6.7 Diffractivity Model 99
6.8 Reflectivity Model 99
6.9 Diffractions 100
6.10 Inverted Model using Diffractions 100

Chapter 1

Introduction

Resolving small scale features using diffractions is now gaining interest in both academic and industrial applications.Traditional Seismic imaging methods use only reflections to create images of subsurface.Diffractions were treated as noise or abandoned stepchildren as referred by Khaidukov [9] and used to suppress in processing and migration steps.But now people have realised the potential of 'the abandoned stepchildren' and they have started adopting it to create high resolution images.

Reflections can only provide you information about the strong reflectors that curve smoothly , it only provides the information at scales larger than the wavelength. We have to resort to diffractions for small scale details or for super resolution. By super resolution we mean higher resolution than the traditional definition of resolution. Typically , the vertical resolution is defined as a quarter of dominant wavelength

$$\lambda = \frac{v}{f} \tag{1.1}$$

where v is the velocity and f is the dominant frequency which varies with both medium and the source properties.

Since our resolution is dependent on wavelength one would question why don't we simply decrease our wavelength to have better resolution images. The reason is decreasing the wavelength of our seismic wave after certain extent is not practical. It is nearly impossible

because of the dissipative nature of overburden which attenuates the high frequences and moreover they are further lost during our standard processing. Diffractions occur when seismic waves interact with objects smaller than their wavelength.Then they must contain the information about those small scale objects like faults , fractures , pinchouts etc. The idea of imaging these signals is to reveal this important information which they carry along to the surface of earth.

Inverting for Diffractions could be a breakthrough as they have huge importance especially for the salt-bodies. Diffractions could help in estimating more accurate velocities at the top of the salt , particulary in rough salt bodies (Sava and Biondi,2004) [10]. Moreover diffractions are the most sensitive velocity information that we have from under the salt. Most of the reflection energy that we record at surface has narrow angles of incidence at the reflector, rendering the analysis of moveout ambiguous.

In this thesis we will discuss two approaches to image or invert the diffractions . The first part of the work is to utilise the difference in behaviour of specular reflection and non-specular diffraction when rays or waves impinges a material discontinuity and the second part of the work is to separate diffractions using a model domain based filter to separate into smooth (reflection part)and rough (diffraction part) model and generate diffractions in Born framework.

Chapter 2

Seismic Diffractions and Imaging

2.0.1 Wave Propagation

Elastic perturbations spatially propagate as waves. The behavior of these waves is determined by the media through which they travel. Seismic imaging studies are primarily concerned with seismic waves traveling through the Earth. The behavior of these waves is unsurprisingly determined by elastic modulii and density of the rocks through which they propagate. These attributes determine the velocity with which the wave propagates. The two types of seismic waves are primary (P) and secondary (S) waves. P-waves transport changes in compression through a medium, and are able to propagate through solids and fluids alike. S-waves transport shear deformation through solid media, and do not propagate through fluids (Chapman, 2004 [11]; Fowler, 2004 [12]). Media properties determine the velocities at which these waves propagate via the following equations:

$$V_p = \sqrt{\frac{\kappa + \frac{4}{3}\mu}{\rho}} \tag{2.1}$$

and

$$V_s = \sqrt{\frac{\mu}{\rho}} \tag{2.2}$$

where κ is the bulk modulus of the media, a measure of incompressibility, μ is the shear modulus, a measure of rigidity, and ρ is the material density (Chapman, 2004 [11]; Fowler,

2004 [12]). The P-wave velocity is always greater than the S-wave velocity. Velocity and density combine to form a quantity known as acoustic impedance, a measure of how much pressure is generated within a medium by vibrations, which is defined in an unperturbed medium by where V_0 and ρ_0 are the velocity and density of an unperturbed medium. Reflectivitycan be defined as the spatial derivative of acoustic impedance (Chapman, 2004 [11]; Fowler, 2004 [12]). Acoustic waves may be reflected or transmitted when they encounter a change in acoustic impedance. The strength of the reflected and transmitted portions of the wave is proportional to the reflectivity of the interface; how large a density and velocity contrast exists there, as well as the incident angle. Fermat's principle states that a wave will take the path through a medium corresponding to a stationary travel time and governs the change in direction that a wave experiences as it passes through an interface (Chapman,2004 [11]; Fowler, 2004 [12]).

2.0.2 Reflections and Diffractions

The interaction of propagating acoustic waves with the media is easier to understand and explain if we examine the context in high frequency approximation or rays. The media is usually characterized as separate regions each with their own material properties (velocities, densities etc) , and with a surface in between where the properties changes instantaneously. The rays describes the waves as particles traveling through space through well defined curves (Cerveny 2005)[13]. The surface with large change in the impedance contrast will give rise to strong reflectivity. A reflected ray will have the same inclination angle with the normal as the incident ray. If the surface has complex geometry the normal vector may change rapidly over short distances, which implies that a small perturbation in reflecting point location will result in a much different reflected vector direction. This can hold true for regions where the reflecting surface has large curvature, or where a vertical discontinuity exists. The incoming parallel rays ,can be described by a ray tube, or a geometric bundle of rays that only have energy flux at the ends, will be scattered

and no longer be parallel to each other after reflection from this irregular surface. This scattering phenomena correspond to diffractions, and diffracted rays may no longer be described by ray tubes containing bundles of parallel rays. The rays now have different directional vectors and any ray tube of diffracted rays will have flux outside describing it's end. Reflection occurs if the radius of curvature of a reflecting surface is greater than that of inbound ray wavelength. For reflecting surfaces with smaller radius of curvature than wavelength, diffraction occurs (Klem-Musatov, 1994 [14]; Khaidukov et al., 2004 [9]). If we consider diffractors as regions which will change rays from neat bundles of ray tubes to scattered non-parallel rays, we find that in addition to reflective regions with irregular surface geometry, caustics which focus inbound rays to single point also create diffractions. Diffractions occur when rays which are initially parallel lose their parallel travel paths (Klem-Musatov, 1994)[14]. Diffractions and reflections are therefore the ends of a continuum of reflective backscattering behavior; no firm boundary between the two exists, and an interaction with a reflecting surface may have both diffractive and reflective components (Khaidukov et al., 2004)[9].

However there are some practical measures and properties based on their interaction with the media which could lead to their separation.The three important properties in which they differ are as follows

1) Diffractions have weaker amplitudes as compared to the reflection events . There magnitudes are one order lower as compared to the reflections for a surface acquistion.In the Figure 2.1 low amplitude hyperbola correspond to diffraction

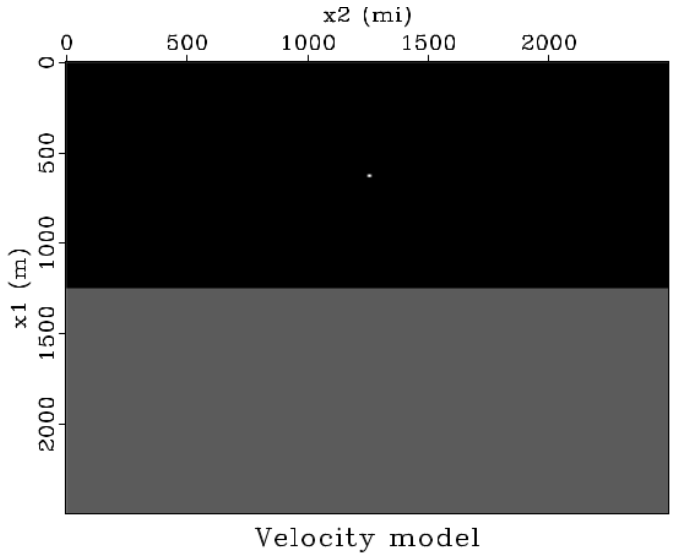


Figure 2.1: Velocity model

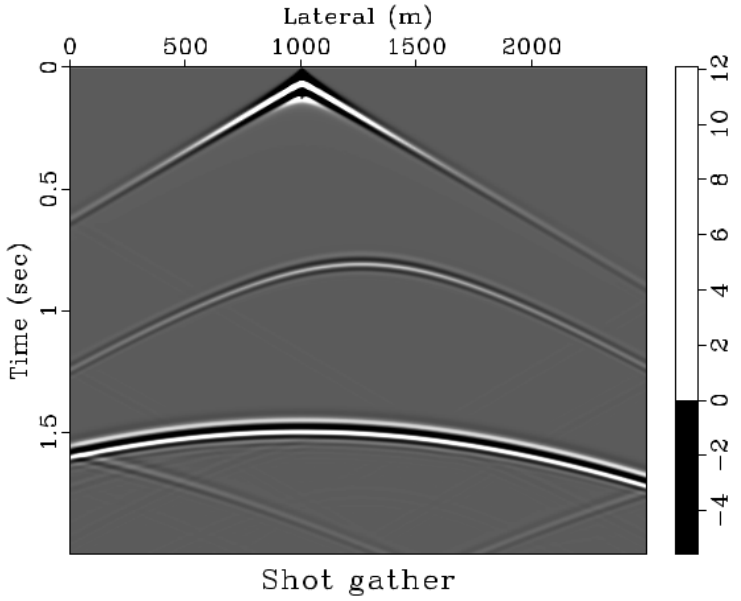


Figure 2.2: Finite difference modeling

2) In zero-offset modeling Diffractions always appear as hyperbolas .Figure 2.3 shows two plane reflectors and one diffractor. In zero-offset kirchhoff modeling the reflectors appears as plane and diffrator appear as hyperbola

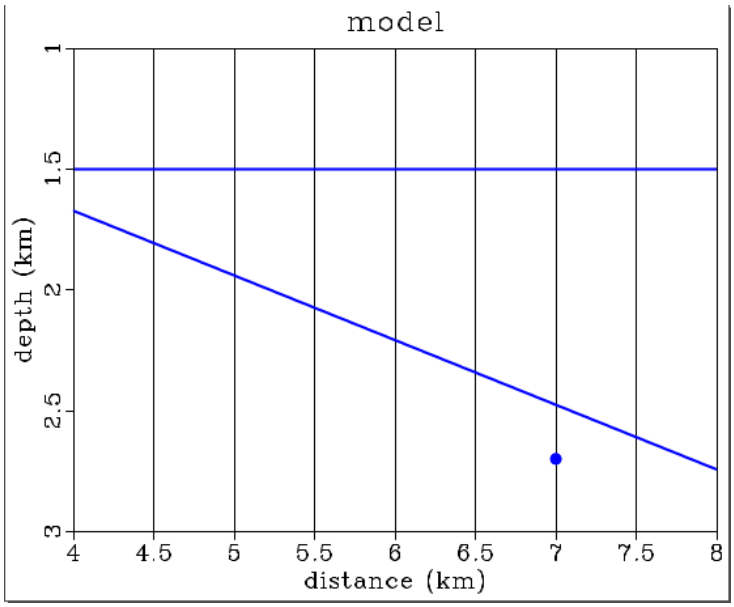


Figure 2.3: Reflectivity Model

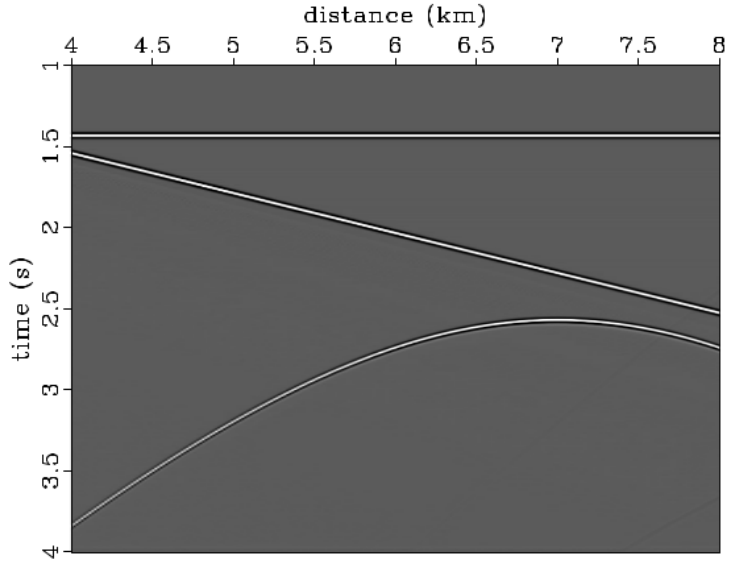


Figure 2.4: Zero Offset Modeling

3) Diffractions doesn't follow snell's law (except edge diffractions).

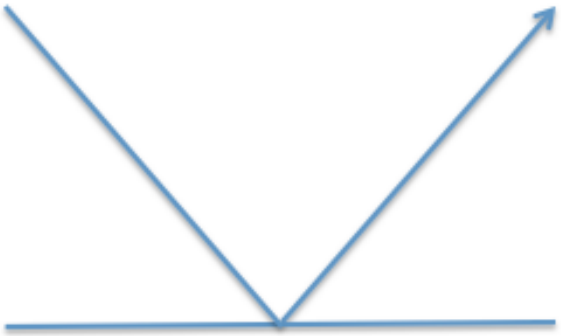


Figure 2.5: Reflections

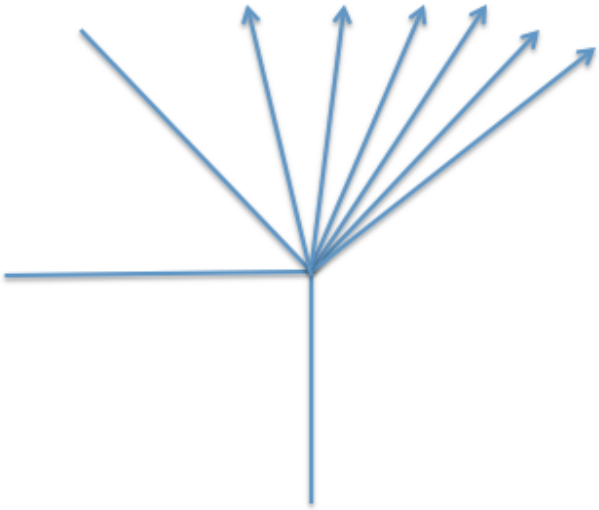


Figure 2.6: Diffractions

2.0.3 Diffraction Imaging and Specularity

The idea of imaging diffractions is not a new concept.The properties that we discussed in the previous section are the most basic one which are exploited by the researchers in different domains to separate and image diffractions. There exist several methods of

diffraction discrimination which uitilizes different criteria in different domains.

An early effort by Harlan et al. (1984) [15] focused on properties of diffractions in the local slant-stack domain. Landa et al. (1987) [16] used differences in travel time properties between diffracted and reflected waves when viewed in common-offset sections to automate diffraction detection. Kanasewich and Phadke (1988) [16]proposed a new form of moveout and amplitude correction to enhance diffraction signal, creating common-fault-point stacked sections. Landa and Keydar (1998)[17]proposed a method of imaging diffractions by focusing energy around common-diffraction points. Fomel (2002) [18] and Fomel et al. (2007) [19] expanded and applied plane-wave destruction filters, first developed by Claerbout (1992) [20], to isolate diffraction from reflection signal in un-migrated common-offset or stacked data. Taner et al. (2006) [21] applied this method in the common-plane-wavedomain. Khaidukov et al. (2004) [9] used the difference between reflection and diffraction moveouts to focus reflections and diffractions at different depths, remove reflections and then de-focus. Landa et al. (2008) [22] and Klokov and Fomel (2012) [23] developed methods for isolating diffractions in migrated dip-angle gathers.

Kozlov et al. (2004)[24] apply a weighting function , based on the Fresnel zone around the specular reflected ray to the Kirchhoff integral to suppress reflections during isotropic migration.Moser and Howard(2008) [25] defined a new weighting function called " Specularity" which measures deviation from Snell's law at reflection point.They defined specualrity as

$$S(s,x,r) = \frac{\left|n^T.T_x\right|}{\|T_x\|} \tag{2.3}$$

where T_x denotes the gradient of $T(s,x,r)$ with respect to x and n the reflector unit normal , depending on x. We will try to utilise this measure directly in Full Waveform Inversion

Chapter 3

Full Waveform Inversion

Seismic Imaging is a parameter estimation problem where the purpose is to reconstruct the parameters (include media properties like density , velocity , attenuation , anisotropy etc) which govern wave propagation . It is often posed as an optimization problem where the purpose is to find the set of model parameters that minimizes a misfit function that quantifies the distance between the observed data and the data predicted by the estimated model. However different misfit function can be designed which will lead to different reconstruction methods based on different imaging conditions. Data domain methods define misfit function based on the characteristics of data . Among them most popular are travel-time tomography , least square migration and full waveform inversion . Alternatively , image domain methods define a misfit function based on the quality of reconstructed images. Full Waveform inversion is a non-linear data fitting procedure which uses entire waveform , unlike in travel time tomgraphy which uses only one seismic arrival , to reconstruct the model parameters. A typical inversion problem consists of two main parts: direct problem (or forward modeling) and inverse problem . In FWI , forward problem consists of solving the wave equation on some domain and inverse problem includes finding optimal parameters which minimizes the misfit function .

3.0.4 Forward Problem

Forward modeling is the heart of seismic imaging. It is a numerical simulation process of actual field surveys which translates the changes in the medium properties in the form of recorded data . Seismologists rely on this data to decipher the important subsurface information. The propagation of seismic waves in the Earth can be modelled using elastic wave equation (or anisotropic wave equation in the most general form)

$$\rho(x)\frac{\partial^2 u(x,t)}{\partial t^2} - \nabla \cdot \sigma = f(x,t) \tag{3.1}$$

that relates the displacement field in the Earth to its mass density ρ , the stress tensor σ and an external force density f . The astounding fact is that it is a linearised version of Newton's second law of motion which balances the momentum of particle $\rho(x)\frac{\partial^2 u(x,t)}{\partial t^2}$ with internal $\nabla \cdot \sigma$ and external forces $f(x,t)$.

However we will limit ourselves to acoustic media approximation and will not discuss the theory of elasticity which is out of the scope of this thesis. In a constant density media , wave equation for the acoustic case can be written as

$$\frac{1}{v(x)^2}\frac{\partial^2 u(x,t)}{\partial t^2} - \nabla^2 u(x,t) = f(x,t) \tag{3.2}$$

where $u(x,t)$ is the pressure wavefield , $f(x,t)$ is the source function and $v(x) = 1/\sqrt{\rho(x)\kappa(x)}$ where ρ is the density , κ is the compressibility , $v(x)$ is the wave velocity.

Analytical solutions to the wave equation exist only for the simple models which donot reflect the structural complexities of the Earth so this deficiency motivatives the development of numerical methods . The numerical methods rely on the discretisation of the continuous seismic wavefields using discrete set of points which can be represented in a form of vector or matrix $u_i(t)$. The differential equations arising from the discretisation of wave equation can be solved in either time domain or frequency domain. Each domain has its own advantages and disadvantages which we will not discuss here and limit our discussions to time domain modeling.However there are ensemble of numerical methods

which can be used to solve wave equation in both domain .Finite difference methods , Pseudospectral methods , Spectral-Element methods , Finite-element methods , Discontinuous Galerkin methods , Direct solutions are the most popular methods but here we tend to discuss only Finite difference method.

3.0.4.1 Finite difference method

Finite difference is the most classy method , frequently applied for numerical solution of wave propagation. It is based on the approximation of exact derivative using neighbouring points. We will demonstrate the finite difference using 1-D wave equation which will helps us to establish basic concepts of numerical stability and grid dispersion which can be easily extended to 2-D and 3-D cases respectively.

The generic way of obtaining the finite-difference approximation of a function is based on truncated Taylor series expansion. Let's say $u(x)$ is the function whose derivative we want to find $\partial_x u(x)$ which is correct upto second order in spacing. We write our derivative as

$$\frac{\partial u(x)}{\partial x} \approx a[u(x+\Delta x) - u(x+\Delta x)] \tag{3.3}$$

a is a scalar coefficient that we seek such that right hand side converges to left hand side as $\Delta x \to 0$. Using Taylor series expansion and cancelling out the even power terms we get

$$\frac{\partial u(x)}{\partial x} \approx a[2\frac{\partial u(x)}{\partial x}\Delta x + \frac{2}{3!}\frac{\partial^3 u(x)}{\partial x^3} + ...] \tag{3.4}$$

If we neglect the higher order terms and choose $a = \frac{1}{2\Delta x}$

$$\frac{\partial u(x)}{\partial x} \approx \frac{1}{2\Delta x}[u(x+\Delta x) - u(x+\Delta x)] + O(\Delta x^2) \tag{3.5}$$

which is a well known approximation (central difference). This scheme can be extended to derive any order Δx approximation such that it involves $2N$ grid points , $x \pm n\Delta x$ with n=1.....N

$$\frac{\partial u(x)}{\partial x} \approx \sum_{n=1}^{N} a_n[u(x+n\Delta x) - u(x-n\Delta x)] \tag{3.6}$$

If expand the terms on the right hand side of the equation using Taylor series , we notice that terms with even powers of Δx will cancel out

$$\frac{\partial u(x)}{\partial x} \approx \sum_{n=1}^{N} a_n \Big[\sum_{k=1}^{\infty} \frac{2}{(2k+1)!} \frac{\partial^{2k+1} u(x)}{\partial x^{2k+1}} (n\Delta x)^{2k+1} \Big] \tag{3.7}$$

Isolating the terms corresponding to k=0 from the integral

$$\begin{aligned} \sum_{n=1}^{N} a_n [u(x+n\Delta x) - u(x-n\Delta x)] = 2\Delta x \Delta_x f(x) \sum_{k=1}^{N} n a_n + \\ \Big[\sum_{k=1}^{\infty} \frac{2}{(2k+1)!} \frac{\partial^{2k+1} u(x)}{\partial x^{2k+1}} a_n (n\Delta x)^{2k+1} \Big] \end{aligned} \tag{3.8}$$

As $\Delta x \to 0$, equating both sides gave , corresponding to k=0 term

$$1 = 2\Delta x \sum_{n=1}^{N} n a_n \tag{3.9}$$

other terms

$$0 = \sum_{n=1}^{N} a_n (n\Delta)^{2k+1} \qquad , k = 1, ..., N-1. \tag{3.10}$$

The resulting finite-difference approximation is then of order $2N$ in Δx

Derivative	Accuracy	−4	−3	−2	−1	0	1	2	3	4
1	2				−1/2	0	1/2			
	4			1/12	−2/3	0	2/3	−1/12		
	6		−1/60	3/20	−3/4	0	3/4	−3/20	1/60	
	8	1/280	−4/105	1/5	−4/5	0	4/5	−1/5	4/105	−1/280

Figure 3.1: Central Difference coefficients for Derivative (source:wikipedia)

In the similar way coefficients for higher order derivatives can be obtained for different order of accuracy. Now let us consider the 1-D wave equation defined on the domain $x = [-L, L]$

$$\frac{1}{v(x)^2} \frac{\partial^2 u(x,t)}{\partial t^2} - \frac{\partial^2 u(x,t)}{\partial x^2} = f(x,t) \tag{3.11}$$

For the sake of simplicity we disregard any external forces , $f(x,t) = 0$ and impose non-zero boundary conditions $u(x,t)|_{t=0} = u_0(x,t)$ and $\dot{u}(x,t) = 0$. The displacements at

boundary $x = L$ and $x = -L$ is zero , $u(x,t)|_{x=L} = u(x,t)|_{x=-L} = 0$. we discretise our domain into equal steps with $2N + 1$ grid points

$$x = i\Delta x \qquad i = -N, ...0..., N \tag{3.12}$$

We use second order approximation of the second derivative obtained from the procedure discussed before

$$\frac{\partial^2 u(x,t)}{\partial t^2} = \frac{1}{\Delta t^2}\Big[u(x,t+\Delta t) - 2u(x,t) + u(x,t-\Delta t)\Big] \tag{3.13}$$

similarly we can write for the w.r.t x . we formulate our wave equation into an explicit time marching scheme .

$$u(x,t+\Delta t) = 2u(x,t) - u(x,t-\Delta t) + \frac{v^2\Delta t^2}{\Delta x^2}[u(x+\Delta x,t) - 2u(x,t) + u(x-\Delta x,t)] \tag{3.14}$$

The solution at next time step can be calculated from previous two values at $t = 0$ and $t = -\Delta t$. This process is repeated till we require . The above strategy is very general and can be repeated using higher order derivatives.

The numerical accuracy of the solution not only depend on the time increment but on grid spacing , finite-difference approximation and even on material properties. The performance of the above solution can be tested if we consider a homogenous media where the analytical solution exist

$$u(x,t) = \frac{1}{2}\Big[u_0(x - vt) + u_0(x + vt)\Big] \tag{3.15}$$

This solution is valid as long it doesn't reach the boundaries . We set the parameters as $v = 5km/s \quad \Delta x = 1km$ and $\Delta t = 0.2s$. The initial wavefield is the derivative of Gaussian field with variance $\sigma = 15km$ result in a dominant wavelength of 30 km.

$$u_0(x,t) = \frac{\partial}{\partial x} e^{\frac{-x^2}{\sigma^2}} \tag{3.16}$$

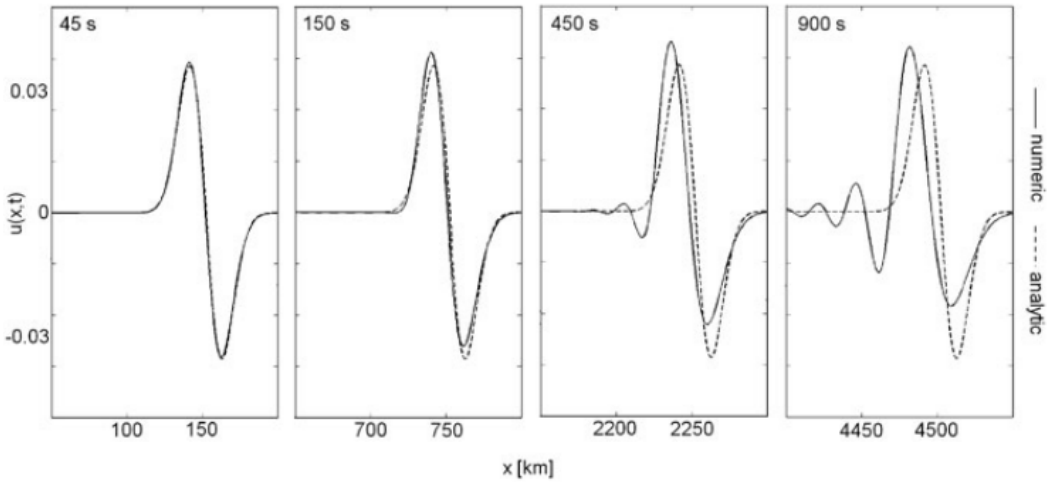


Figure 3.2: Numerical vs Analytical solution 1-D wave equation (source:Fichtner [1])

In the Figure 3.2As you can see that after 450 s the numerical solution tends to be dispersive , meaning high frequencies seems to propagate slower, even if the analytical solution is non-dispersive. This undesirable phenomena is referred as *numerical or grid dispersion* . In the next section we will discuss it in detail and provide measures to diminish it.

3.0.4.2 Stability Analysis

Von Neumann analysis is a powerful tool to study the stability anlysis that we observed in previous section. We will assume a homogenous medium defined on the interval $[-\pi, \pi]$ where we can represent our solution in the form of Fourier series.

$$u(x,t) = \sum_{k=1}^{k=\infty} \phi_k(t) e^{ikx} \tag{3.17}$$

In discrete form our wavefield solution can be written as

$$u(x,t) = \sum_{k=-N/2}^{k=N/2} \phi_k(t) e^{ikp\Delta x} \tag{3.18}$$

Please note that the above discretisation $\Delta x = 2\pi/N$ limits the range to nyquist wave number $\pi/\Delta x$, $k \in [-\pi/\Delta x, \pi/\Delta x]$. Since we assume no force function in our example , we insert the above expression into semi-discrete wave equation.

$$\frac{\partial^2 u}{\partial t^2} = \frac{v^2}{4\Delta x^2}\left[e^{2ik\Delta x} - 2 + e^{-2ik\Delta x}\right]\phi_k(t)e^{ikp\Delta x} \tag{3.19}$$

$$= \frac{-v^2}{\Delta x^2}\sin^2(k\Delta x)\phi_k(t)e^{ikp\Delta x} \tag{3.20}$$

Using second order finite difference in time and substituting the Fourier series

$$\phi_k(t+\Delta t) - 2\phi_k(t) + \phi_k(t-\Delta t) = \frac{-v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x)\phi_k(t) \tag{3.21}$$

We introduce a new variable ψ_k to remove explicit dependence of ϕ_k at time $t - \Delta t$

$$\psi_k(t+\Delta t) = \phi_k(t+\Delta t) - \phi_k(t) \tag{3.22}$$

After substitution we get

$$\psi_k(t+\Delta t) - \psi_k(t) = \frac{-v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x)\phi_k(t) \tag{3.23}$$

Using the above two equations we can form a linear system independent of time $t - \Delta t$.

$$\begin{pmatrix} \psi_k(t+\Delta t) \\ \phi_k(t+\Delta t) \end{pmatrix} = \begin{pmatrix} 1 & -\frac{-v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x) \\ 1 & 1 - \frac{-v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x) \end{pmatrix} \begin{pmatrix} \psi_k(t) \\ \phi_k(t) \end{pmatrix} \tag{3.24}$$

We can advance Fourier coefficients in time by the application of this matrix ,

$$\begin{pmatrix} \psi_k(j\Delta t) \\ \phi_k(j\Delta t) \end{pmatrix} = M_k^j \begin{pmatrix} \psi_k(0) \\ \phi_k(0) \end{pmatrix} \tag{3.25}$$

where

$$M_k = \begin{pmatrix} 1 & -\frac{-v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x) \\ 1 & 1 - \frac{-v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x) \end{pmatrix} \tag{3.26}$$

Now all the relevant properties of the algorithm especially stability and dispersion is encapsulated in the form eigen values and vectors of the above linear system. For this algorithm to be stable the absolute of eigen values λ_k must be less than 1 , otherwise it will grow indefinitely . In order to faciliate eigen value analysis , we define certain parameters for neatness of the solution .

$$\gamma_k = \sin^2(k\Delta x) \tag{3.27}$$

and $Courant \quad Number$

$$c = \frac{v\Delta t}{\Delta x} \tag{3.28}$$

In terms of above parameters , the eigen values can be written as

$$\lambda_k = 1 - \frac{1}{2}c^2\gamma^2 \pm \sqrt{\frac{1}{4}c^4\gamma^4 - c^2\gamma^2} \tag{3.29}$$

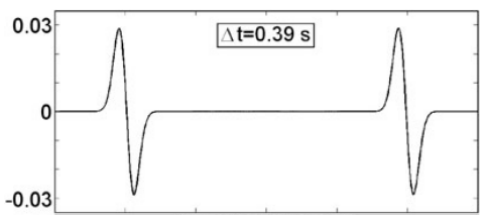
If term under square root is real then we can have $\lambda_k > 1$. To ensure stablity the term inside the square root should be negative ($c^2\gamma^2 < 4$) such that λ_k is imaginary , then

$$|\lambda_k| = \sqrt{\lambda_k \lambda_k^*} = 1 \tag{3.30}$$

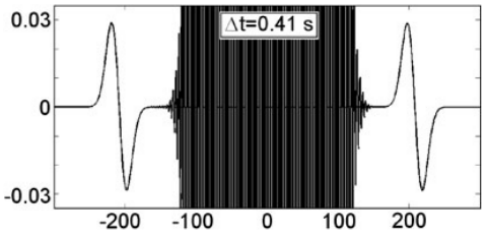
Since γ ranges between 0 and 1 so ,

$$c = \frac{v\Delta t}{\Delta x} < 2 \tag{3.31}$$

The above condition ensures the stability to the solution of 1-D wave equation using second-order finite difference equation . This condition is known as CFL condition named after R. Courant, K. Friedrichs and H. Lewy (Courant et al.1928) . With $\Delta x = 1km$ and $v = 5km/s$, the CFL condition imposes a maximum time increment of 0.4 s. Smaller time increments are predicted to yield stable solutions, whereas larger time increments are expected to lead to solutions that grow indefinitely as the iteration proceeds. The prediction of the CFL condition is remarkably accurate. As you can see in the Figure **??** the solution is stable for $\Delta t = 0.39s$, but it 'explodes' for $\Delta t = 0.41s$.



(a) Numerical solution for $\Delta t = 0.39s$



(b) Numerical solution for $\Delta t = 0.41s$

In general CFL condition have the form

$$\Delta t < const * \frac{min(\Delta h)}{max(v)} \tag{3.32}$$

It is possible to circumvent the CFL condition using implicit schemes but there computational cost compensates the advantages.

In order to quantify numerical dispersion , we assume the initial state $(\psi_k(0), \phi_k(0))$ is the eigen vector of the matrix M_k.The iteration that advances the Fourier coefficients in time then simplifies to

$$\begin{pmatrix} \psi_k(j\Delta t) \\ \phi_k(j\Delta t) \end{pmatrix} = \lambda_k^j \begin{pmatrix} \psi_k(0) \\ \phi_k(0) \end{pmatrix} \tag{3.33}$$

As λ_k is a complex number it can written in Euler form

$$\lambda_k = |\lambda_k| e^{i\alpha_k} \tag{3.34}$$

we write α_k as ,

$$\alpha_k = \Delta t k \bar{v} \tag{3.35}$$

Introducing the above expression to Fourier representation we get

$$u(j\Delta t) = \sum_{k=-N/2}^{k=N/2} \phi_k(0) |\lambda_k| e^{ik(j\Delta t\bar{v} + p\Delta x)} \tag{3.36}$$

equation reveals that $\bar{v}$ plays role of numerical wavespeed. If the numerical solution was exact it would have been equal to exact wavespeed and independent of wave number k. As a result of the discretisation,$\bar{v}$ depends upon wavenumber thus our numerical solution is dispersive in contrast to exact solution

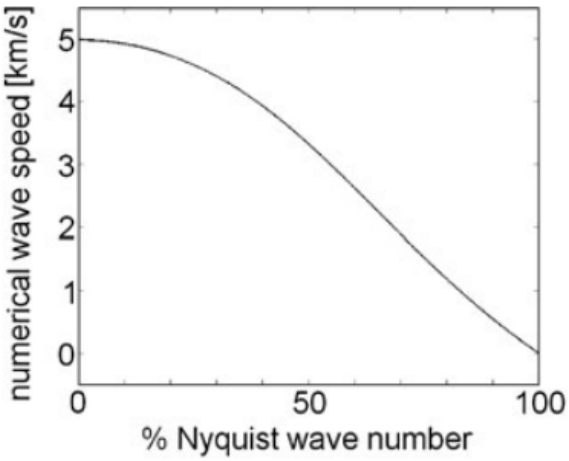


Figure 3.4: Numerical Dispersion for exact wavespeed $v = 5km/s$ (source:Fichtner [1])

The components with wave number close to zero propagate at the correct speed, v = 5 km/s. As the wave numbers approach the Nyquist wave number,the propagation speed tends to zero. Numerical dispersion is, of course, not a special property of our particular example. All numerical solutions are dispersive to some degree, regardless of the method used.

3.0.4.3 Point source implementation

In seismic surveys , we use explosives or air-gun as our sources.These physical sources are best described as point sources . Point sources has the form

$$f(x,t) = \delta(x - x_0)\delta(t - t_0) \tag{3.37}$$

It is not feasible to implement delta function in time domain and we usually implement its band limited definition in time domain

In the frequency domain , the delta function can be written as

$$\delta(\omega) = \frac{1}{\sqrt{2\pi}} e^{-i\omega t} \tag{3.38}$$

As the result of the discretisation , ω is limited to the Nyquist frequency ω_{max} which we can resolve.The optimal band limited version of δ is

$$\bar{\delta}(\omega) = \frac{1}{\sqrt{2\pi}}[H(\omega + \omega_{max}) - H(\omega - \omega_{max})] \tag{3.39}$$

or

$$\bar{\delta}(t) = \frac{\omega_{max}}{\pi} sinc(\omega_{max}(t - t_0)) \tag{3.40}$$

A discrete band limited δ function can be found by sampling at discrete time sampling points $t_n = i\Delta t$.

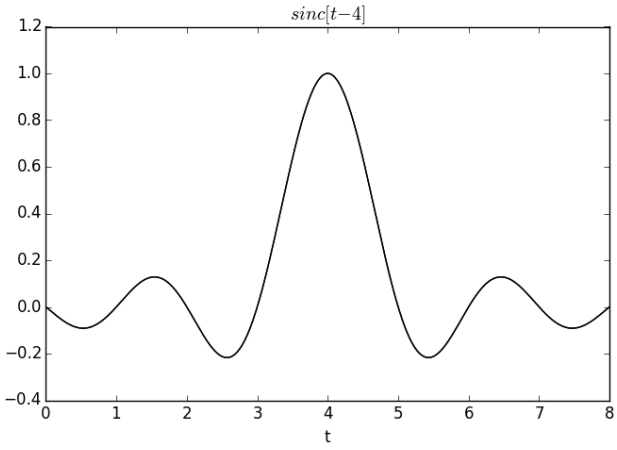


Figure 3.5: $sinc[t - 4.0]$

However our source is usually impulsive which means it remains active only for a small amount of time so to restrict our source function to small period of time we can employ windowing functions . Possible windows are Gaussian window or Kaiser windows. The most popular source used in Seismic function is Ricker wavelet source which like a windowed version of *sinc* function.

$$w(t) = (1 - 2\pi^2\nu_0^2(t - t_0)^2)e^{-\pi^2\nu_0^2(t-t_0)^2} \tag{3.41}$$

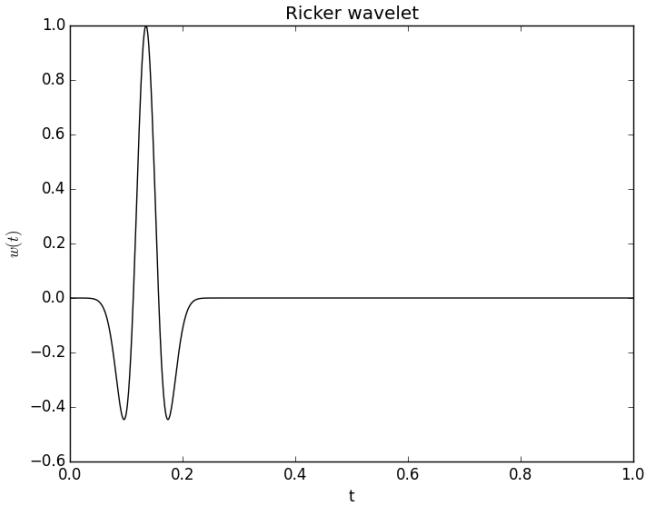


Figure 3.6: $Ricker, \nu_0 = 10Hz$

3.0.4.4 Boundary Conditions

It is not possible to solve for the whole physical domain on the computer machines so usually we have to limit our computational domain to some part of the physical domain. The biggest problem while restricting is it introduces reflections from these artificial boundaries which don't exist in the Earth. If we do not treat them adequately , they will pollute our numerical solution and error will be dominated so we have to introduce some artificial boundaries which will absorb these unwanted reflection and prevent decading our solution. This introduction of artificial boundaries are known as Absorbing Boundary conditions in literature. There are numerous ABC's based on the paraxial approximation of wave equation (e.g. Engquist and Majda, 1977 [26]). Their efficiency depends strongly on the angle of incidence. The absorption is perfect only for waves propagating normal to the boundary. We will discuss only Paraxial approximation of wave equation boundary conditions and sponge boundary conditions . We will also brief about the more extensive boundary conditions known as Perfectly Matched Layer but we will not derive the regarding equations here.

3.0.4.4.1 Paraxial approximation ABC's : The idea of these boundary conditions is when an incident wave hits the boundary the reflected wave direction will be approximately opposite so this idea lead to paraxial approximation of wave equation. The acoustic wave equation in a full space is

$$\frac{\partial u^2}{\partial t^2} - \Big[\frac{\partial u^2}{\partial x^2} + \frac{\partial u^2}{\partial y^2} + \frac{\partial u^2}{\partial z^2}\Big] = 0 \tag{3.42}$$

The plane wave solution is written as

$$u = e^{i(k.x - \omega t)} \tag{3.43}$$

The wavenumber vector $k = [k_x, k_y, k_z]$ is related to wavespeed v and angular frequency ω via dispersion relation.

$$\frac{\omega^2}{v^2} = k_x^2 + k_y^2 + k_z^2 \tag{3.44}$$

or

$$k_z^2 = \pm\sqrt{\frac{\omega^2}{v^2} - (k_x^2 + k_y^2)} \tag{3.45}$$

Taking only positive sign into consideration as it will denote increasing in z direction. The resulting wave equation is

$$\Big[\frac{\partial u}{\partial z} - i\sqrt{\frac{\omega^2}{v^2} - (k_x^2 + k_y^2)}\Big] u(k_x, k_y, z, \omega) = 0 \tag{3.46}$$

It is a wave equation in mixed space-frequency-wavenumber domain obtained impractical from computational point of view . We expand the term in square using truncated Taylor series expansion upto second order however higher order approximations can also be derived.

$$\frac{vk_z}{\omega} = 1 + O\Big[\Big|\frac{v^2}{\omega^2}(k_x^2 + k_y^2)\Big|\Big] \tag{3.47}$$

$$\frac{vk_z}{\omega} = 1 - \frac{1}{2}\frac{v^2}{\omega^2}(k_x^2 + k_y^2) + O\Big[\Big|\frac{v^2}{\omega^2}(k_x^2 + k_y^2)\Big|^2\Big] \tag{3.48}$$

Following Clayton and Engquist [26](1977), the approximations (51) and (52) are commonly referred to as the A1 and A2 paraxial approximations.The corresponding one-way wave equations are :

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial z} = 0 \tag{3.49}$$

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial z} - \frac{1}{2}\Big(\frac{\partial u^2}{\partial x^2} + \frac{\partial u^2}{\partial y^2}\Big)u = 0 \tag{3.50}$$

The second equation is well known as 15^0 approximation of wave equation however we can assess the absorption efficiency by computing the reflection coefficient R that determines the amplitude of the reflected waves from these artificial boundaries. We write our solution as

$$u = u_{z+} + Ru_{z-} \tag{3.51}$$

where as $u_{z+} = e^{i(k_x x + k_y y + k_z z - \omega t)}$ and $u_{z-} = e^{i(k_x x + k_y y - k_z z - \omega t)}$

$$k = \frac{\omega}{v} \begin{pmatrix} \cos\phi \sin\theta \\ \sin\phi \sin\theta \\ \cos\theta \end{pmatrix} \tag{3.52}$$

where $\phi \in [0, 2\pi]$ is the angle between k and positive x-axis and $\theta \in [0, \pi/2]$ is the incidence angle between k and positive z-axis.

For A1 condition

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial z} = 0 \tag{3.53}$$

we get

$$R = -\frac{1 - \cos\theta}{1 + \cos\theta} \tag{3.54}$$

For A-2 condition

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial z} - \frac{1}{2}\Big(\frac{\partial u^2}{\partial x^2} + \frac{\partial u^2}{\partial y^2}\Big)u = 0 \tag{3.55}$$

we get

$$R = -\frac{(1 - \cos\theta)^2}{(1 + \cos\theta)^2} \tag{3.56}$$

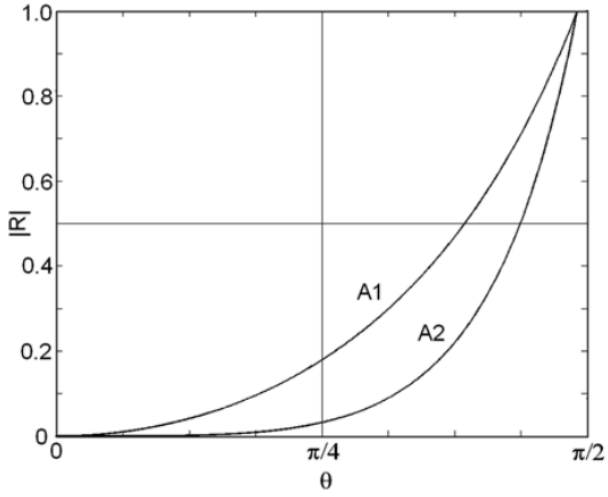


Figure 3.7: Reflection coefficients vs Incidence angle (source:Fichtner [1])

From the plot between (Figure 3.7)reflection coefficients and angle of incidence , one can see that both boundary conditions are perfectly absorbing for plane waves parallel to

z-axis . The reflection coefficient for A1-condition grows more rapidly but for incidence angle more than $\pi/4$ both results in unacceptedly large reflections therefore higher order boundary conditions could be built using the same procedure (including more terms from truncated Taylor expansion)

3.0.4.4.2 Sponge Absorbing Boundary conditions : Sponge ABC's were proposed by Cerjan et al (1985)[27] . The idea of sponge is to introduce sponge like layers around the computational domain where reflections attenuates exponentially . We extend the model domain with layers of certain thickness , usually thickness varies from 20-60 .We solve the wave equation in this extended domain and in this artificial layers we multiply the wavefield with an exponentially decaying term. We commonly use the factor $d(u)$

$$d(u) = \exp(-[0.015 * (nb - i)]^2) \tag{3.57}$$

where nb is the layer thickness . These boundary conditions are the simplest one and found to be more effective than 15^0 ABC.

3.0.4.4.3 Perfectly Matched Layer :Sponge Boundary conditions proves to be effective with finite difference method but it requires unacceptably large sponge layers with higher order methods such as spectral method and even in higher order approximations in FD , nb goes beyond 60 .A more efficient boundary layer technique was introduced by Bérenger (1994) [28], who proposed to modify the electrodynamic wave equation inside a perfectly matched layer (PML) such that the solutions decay exponentially without producing reflections from the boundary between the regular medium and the PML. Since Bérenger's original invention, the PML method has been extended to a large variety of wave propagation phenomena . There are various version of PML available (Split PML , Convolutional PML etc)and they form a vast body of literature. The idea is to modify the wave equation inherently such that the incident waves decay rapidly .It is modified such that it still matches the wave equation and no reflections are produced inside the

PML region while transmission. However for further details , the reader is referred to (Carcione et al., 2002) [29] for Split PML , (Roden and Gedney, 2000) [30]for Non-Split Convolutional PML , (Festa et al., 2005) [31] for Convolutional PML .

3.0.5 Inverse Problem

The objective in full waveform inversion is to find an optimal model $\bar{m}$ representing Earth that minimize the misfit functional χ, to quantify the difference between the observed data and predicted data.A model comprises the spatial distribution of media properties like P-wavespeeds $v_p(x)$, S-wavespeeds $v_s(x)$, density $\rho(x)$ etc.

Since χ is a non-linear functional of m , we approximate the optimal model $\bar{m}$ with iterative schemes starting with an initial model m_0 .The iterative scheme updates model to a new model that better expalins the data than its predecessor:

$$m_{i+1} = m_i + \alpha_i d_i \qquad with \quad \chi(m_{i+1}) < \chi(m_i) \tag{3.58}$$

The choice of descent direction d_i , and the step length $\alpha_i > 0$ depends upon the choice of minimization scheme used to approximate $\bar{m}$. The iterative procedure is repeated till our model satisifies our data sufficiently well. We will brief about the conditions that $barm$ needs to satisfy to consider to be minimum of χ and then we will review the optimization methods and regularization schemes that ensures the covergence towards physically meaningful models.

3.0.5.1 Mimima , Convexity and Non-Uniqueness

Full Waveform Inversion involves the computationally expensive forward modeling and a high-dimensional model space that render the application of probabilistic schemes impractical. Thus we are limited to the deterministic solution of the inverse problem which means to find the model $\bar{m}$ that minimizes the misfit functional χ

We consider a neibhourhood $N_r(\bar{m})$ of radius r around the optimal model $\bar{m}$ and define

our model space as Ω which include all the permissible Earth models.

$$N_r(\bar{m}) = m \in \Omega \quad |||m - \bar{m}||_2 < r \tag{3.59}$$

The model $\bar{m} \in \Omega$ is a local minima of χ if there exists a small neibhourhood where $\chi(\bar{m})$ is smalller than $\chi(m)$

If it is smaller for all the models m in the neighbourhood then it is a strict local minima . For a global minima the $\chi(\bar{m})$ is smaller than or equal to $\chi(m)$ for all the models that belongs to our domain not for a particular neighbourhood. For a strict global minima there is a unique model for whom $\chi(\bar{m}) < \chi(m)$ for all $m \in \Omega$. The distinction between minima and strict minima is important which relates to uniqueness of the solution. If the minima is strict it means there is unique solution otherwise the solution is non-unique.

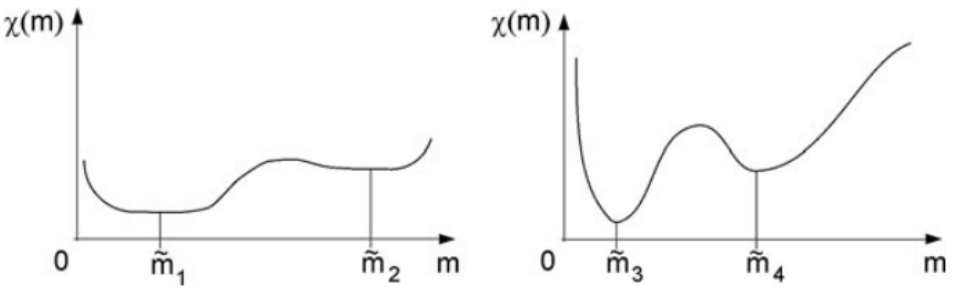


Figure 3.8: Left:m_1 is global minima and m_2 is local minima Right: m_3 is strict global minima and m_4 is strict local minima (source : Fichtner [1])

The existence of global minima invloves the approximation in physics of the problem that were used to model . The locality and uniqueness depends upon the convexity properties of the model space Ω . We define a model space convex if the line joining two models also fall in the model space.

$$\alpha m_1 + (1 - \alpha) m_2 \in \Omega \tag{3.60}$$

for all $\alpha = [0, 1]$ and for all m_1 and m_2 in Ω

We can define convex misfit functionals provided that our model space is covex . We

will evaluate the conditions in form of convexity with which one can ensure that a global minima exist . The convex misfit functionals are defined as

$$\chi(\alpha m_1 + (1-\alpha)m_2) \leq \alpha\chi(m_1) + (1-\alpha)\chi(m_2) \tag{3.61}$$

The convexity of the misfit functional ensures that the local minima is actually a global minima but it doesn't ensures the uniqueness .

Proof 1: Global minima ensureness Let's assume that m' is local minima of χ . There exists a neighbourhood $N_r(m')$ around m' such that

$$\chi(m') \leq \chi(m) \tag{3.62}$$

for all $m \in N_r(m')$. The convexity implies that any line $m(\alpha)$ connecting m' to any other model m'' in the model space Ω is still in Ω.

$$m(\alpha) = \alpha m'' + (1-\alpha)m' = m' + \alpha(m'' - m') \in \Omega \tag{3.63}$$

for all $\alpha \in [0, 1]$

By choosing $\alpha \leq r||m'' - m'||_2^{-1} = \alpha_r$ we ensure $y(\alpha_r)$ is also in $N_r(m')$. Therefore

$$\chi(m') \leq \chi(m(\alpha_r)) = \chi(\alpha m'' + (1-\alpha)m') \leq \alpha\chi(m'') + (1-\alpha)\chi(m') \tag{3.64}$$

Rearranging equations on both sides we get

$$\chi(m') \leq \chi(m'') \tag{3.65}$$

Since the choice of m'' is arbitary so the relation implies that convexity is sufficient to guarantee that a local minima is infact global minima.

Proof 2: Uniqueness The strict convexity $\chi(m') < \chi(m)$ ensures the global minima m' is unique. Let us consider there exists another minima m'' of χ such that $m' \neq m''$ so $\chi(m') = \chi(m'')$. The mid-point of the models $(m' + m'')/2$ should also lie in the model space Ω as per the convexity.

$$\chi(m') < \chi((m' + m'')/2) < \chi(m'/2) + \chi(m''/2) \tag{3.66}$$

Rearranging it follows that $\chi(m') < \chi(m'')$ which contradicts the assumption that m'' is also a minima of χ

It shows that strict convex misfit functional ensures the uniqueness of global minima so we would like our misfit functional to be convex . However it is not handy to work on this strict definition of convex functional as we can't evaluate the misfit along the line between two models in our model space.

The more handy is to develop the criteria in terms of derivatives .If χ is convex , for any two models m_1 and m_2

$$\chi[m_1 + \alpha(m_2 - m_1)] \leq \chi[m_1] + \alpha[\chi(m_2) - \chi(m_1)] \tag{3.67}$$

Rearranging we get

$$\frac{1}{\alpha}(\chi[m_1 + \alpha(m_2 - m_1)] - \chi[m_1]) \leq \chi(m_2) - \chi(m_1) \tag{3.68}$$

In the limit $\alpha \to 0$

$$\chi[m_2] \geq \chi[m_1] + (m_2 - m_1)\nabla_m \chi(m_1) \tag{3.69}$$

Exapnding $\chi(m_2)$ using Taylor series expansion, we get

$$\chi(m_2) = \chi(m_1) + (m_2 - m_1)\nabla_m \chi(m_1) + \frac{1}{2}(m_2 - m_1).H_\chi(m_1).(m_2 - m_1) + O(|m_2 - m_1|)^3 \tag{3.70}$$

Since equation 74 holds for any arbitary m_2 , we get

$$(m_2 - m_1).H_\chi(m_1).(m_2 - m_1) \geq 0 \tag{3.71}$$

It means that Hessian of a strictly convex function should be positive definite. It could be established without proof that for a misfit functional to be convex , it's Hessian should be positive definite. This condition still could be verified as compared to the original definition of convex functionals. But however computing Hessian for a problem at large scale is still not considered practical .

3.0.5.2 Optimization methods

Descent methods are the most popular methods used for solving the optimization problem iteratively.The general idea of descent methods is to start with an initial model m_0 , which is usually our best guess before inversion , and updating the model iteratively in a particular descent direction d_0 over the step length α_0.

$$m_1 = m_0 + \alpha_0 d_0, \quad \chi(m_1) < \chi(m_0) \tag{3.72}$$

The necessary condition for the construction of descent direction is

$$\chi(m_1) = \chi(m_0 + \alpha_0 d_0) < \chi(m_0) \tag{3.73}$$

If we consider the step length α_0 to be very small , then using the limits or truncated Taylor series upto first order, we get

$$d_0 . \nabla_m \chi(m_0) < 0 \tag{3.74}$$

The descent direction has to satisfy the above condition . We choose $d_0 = -\nabla_m \chi$ which will ensures the above condition.

$$-\left[\nabla_m \chi(m_0)\right]^2 < 0 \tag{3.75}$$

However other descent directions are also possible which lead to different descent methods. Another possible descent direction is

$$d_0 = -A . \nabla_m \chi(m_0) \tag{3.76}$$

only if A is a positive definite matrix , because then

$$h_0 . \nabla_m \chi(m_0) = -\nabla_m \chi(m_0) . A . \nabla_m \chi(m_0) < 0 \tag{3.77}$$

The optimal step length should be such that our misfit functional should be minimum . $\chi(m_{i+1})$=$\chi(m_i + \alpha_i d_i)$ is minimum.

$$\frac{\partial \chi(m_i + \alpha_i d_i)}{\partial \alpha_i} = 0 \tag{3.78}$$

Evaluating the above condition and using the linear approximation , we get

$$\alpha_i = -\frac{d_i.\nabla_m \chi(m_i)}{d_i.H_\chi(m_i).d_i} \tag{3.79}$$

The other alternative to avoid the computation of Hessian to use a line search method like back tracking , parabolic line search method etc.

3.0.5.2.1 Steepest Descent Method In steepest descent method we find descent direction d_i such that $||d_i||_2 = 1$. The choice of descent direction for steepest descent is

$$d_i = -\frac{\nabla_m \chi(m_i)}{||\nabla_m \chi(m_i)||_2} \tag{3.80}$$

The step length could be calculated using either line search or the analytical method . However being conceptually simple , it is rarely used because of its slow convergence.

3.0.5.2.2 Conjugate Gradient Method The conjugate gradient involves finding the descent directions that are conjugate to each other.To illustrate we start with quadratic misfit functionals.

$$\chi(m) = \chi(m') + \frac{1}{2}(m - m').H_\chi(m).(m - m') \tag{3.81}$$

The Hessian is assumed to be positive definite , so the m' is a unique global minima of misfit functional. The conditions for the directions to conjugate is as follows .

$$d_i.\nabla_m \chi(m_j) = 0 \tag{3.82}$$

$$d_i.H_\chi(m).d_j = 0 \tag{3.83}$$

By ensuring the conjugacy of descent directions , it will ensure the solution in limited number of iterations.

$$d_k = \begin{cases} -\nabla\chi(m_0) & k = 0 \\ -\nabla\chi(m_k) + \beta_k d_{k-1}, & k \geq 1 \end{cases} \tag{3.84}$$

There numerous ways to compute β_k , we use a hybrid scheme combining Hestness-Stiefel and Dai-Yuan schemes:

$$\beta_k = max(0, min(\beta_k^{HS}, \beta_k^{DY})) \tag{3.85}$$

where

$$\beta_k^{HS} = \frac{< \nabla\chi(m_k), \nabla\chi(m_k) - \nabla\chi(m_{k-1}) >}{< d_{k-1}, \nabla\chi(m_k) - \nabla\chi(m_{k-1}) >} \tag{3.86}$$

$$\beta_k^{DY} = \frac{< \nabla\chi(m_k), \nabla\chi(m_k) >}{< d_{k-1}, \nabla\chi(m_k) - \nabla\chi(m_{k-1}) >} \tag{3.87}$$

3.0.6 Elements of practical FWI

In previous sections we discussed the various elements involved in Full waveform inversion in detail . In this section we will put the information at the right place to get the real picture. As described previously , FWI consists of solving both forward problem and inverse problem in an iterative manner. In time domain , acoustic wave equation for isotropic and constant density media can be written as

$$\Big(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\Big)u(x,t) = s(x,t) \tag{3.88}$$

where v is the wavespeed , $u(x,t)$ is the pressure wavefield and $s(x,t)$ is the source function. In compact form it can be written as

$$Au = s \tag{3.89}$$

where

$$A(x,t,v) = \Big(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\Big) \tag{3.90}$$

The inverse problem consists in finding optimal model that minimizes the misfit functional χ. The misfit functional is a function of observed data d_{obs} and predicted data d_{pre} at receiver positions. The common choices for objective functional invloves cross-correlations , norms and deconvolutional type objective functions or combination of them. We employ L_2 norm type objective function however it is not the best type of objective function but

it is relatively simple to use.

$$min(\chi) = min\frac{1}{2}\int_{t=0}^{t=T}\sum_{s=1}^{s=N_s}\sum_{r=1}^{s=N_r}||d_{obs}(x,t) - d_{pre}(x,t)||_2^2 \tag{3.91}$$

where N_s is the number of sources , N_r is the number of receivers , d_{obs} is the observed data and d_{pre} is predicted data using forward simulation.

The inverse problem is solved as an optimization problem , the updated model at n^th iteration m_{n+1} is

$$m_{n+1} = m_n + \alpha\Delta m \tag{3.92}$$

$$\Delta m = \beta\nabla_m\chi_n \tag{3.93}$$

For Newton methods the β is the inverse of Hessian

3.0.6.1 The Gradient: Computaion and Interpretation

For any optimization problem , finding the optimal model requires the derivative of misfit functional χ . d_{obs} is the wavefield recorded at the receiver position . We write $d_{obs} = Pu$, where P is the projection operator from model space to receiver space and u is the complete wavefield

$$\frac{\partial\chi}{\partial v} = g = \int_{t=0}^{t=T}\sum_{s=1}^{s=N_s}\sum_{r=1}^{s=N_r}\Big(P\frac{\partial u(x,t)}{\partial v}\Big)^{\dagger}(Pu(x,t) - d_{obs}(x,t)) \tag{3.94}$$

We need to compute $\frac{\partial u}{\partial v}$. It can't be done analytically as we don't have an analytical solution of wave equation for complex media . In the most general case it can done numerically using finite difference method or taking the derivative of the forward problem $(Au = s)$.

$$\frac{\partial u}{\partial v} = \begin{cases} \frac{\partial u}{\partial v} = \frac{u(x,t,v+\Delta v) - u(x,t,v)}{\Delta v} \\ A\frac{\partial u}{\partial v} = -\frac{\partial A}{\partial v}u(x,t,v) \end{cases} \tag{3.95}$$

The second one is more numerically accurate as it involves solution of $N \times N_s$ where N is the total number of discrete points $nx \times nz$. Physically $\frac{\partial u}{\partial v}$ it represents a diffracted wavefield by a heterogeneity at model position i .

Alternatively the gradient can be computed using adjoint state method which requires only two forward problems per source $2 \times N_s$.

The exact derivation of the gradient can be found in appendix and the final relation comes out to be.

$$\frac{\partial \chi}{\partial v} = \int_{t=0}^{t=T} R\Big\{ \frac{2}{v^3} * \frac{\partial^2 u}{\partial t^2} * \lambda(x,t) \Big\} \tag{3.96}$$

The gradient expression is proportional to cross-correlation of forward wavefield $u(x,t)$ and back-propagated wavefield $\lambda(x,t)$. The real part ensures the model parameters have physical meaning. The above relation from adjoint state method comes out be simple because our wave propagation operator is self adjoint ($A = A^\dagger$)

3.0.6.1.1 Preconditioner : The descent direction can be improved by using a preconditioner. There is no general method to compute the preconditioner however it should not be much expensive.For gradient based methods the left preconditioner has the form

$$\Delta m = P \nabla_m \chi \tag{3.97}$$

where P is the preconditioner . For Newton and its variant methods the preconditioner is the inverse of the Hessian however it is to much expensive to compute the exact Hessian as it require the computation of diffracted wavefields ($\frac{\partial u}{\partial v}$ or Frechet derivative.

A good choice of preconditioner to use is to compute inverse of pseudo Hessian which was introduced by Shin . He approximated the diffracted wavefields as

$$\frac{\partial u}{\partial v} = -\frac{\partial A}{\partial v} u \tag{3.98}$$

as if the modeling operator A is behaving as identity .It can be calculated just cross-correlating the adjoint sources.The resulting pseudo Hessian is

$$H^{ps} = \Big(\frac{\partial A}{\partial v}\Big)^\dagger \Big(\frac{\partial A}{\partial v}\Big) \tag{3.99}$$

Since the approximation of inverse of Hessian is desired , only diagonal terms are taken into account

$$P = diag\Big(\frac{1}{H^{ps} + \beta}\Big) \tag{3.100}$$

3.0.6.1.2 Physical Interpretation :We refer to the imaging condition as the operator that goes from the data space to the model space, and determines which parameters in the model space must be modified in order to explain the measured data. The gradient gives the imaging condition for FWI.The gradient expression illustrates that the correlation of the direct and back-propagated wavefield will be non-zero on all the points along an ellipse (Figure 3.9).

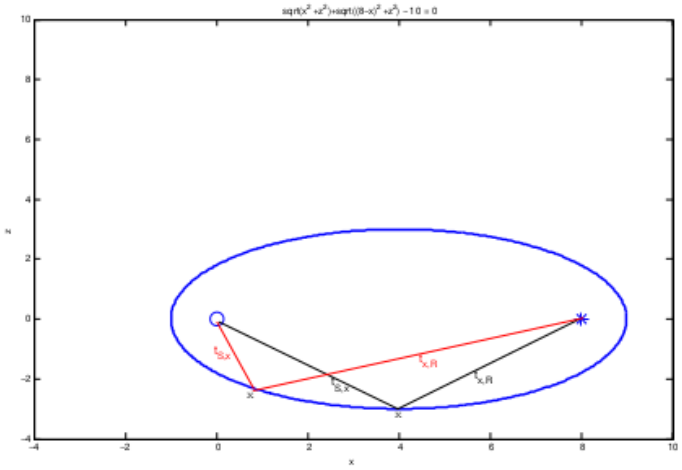


Figure 3.9: Example of the points satisfying the imaging condition $t_{S,x} + t_{x,R} = T$, for a reflection residual at time T = 5s, in a model with v = 2m/s, with a source at $x_S = 0$ and a receiver at $x_R = 8$ m, resulting in an offset of D = 8m. The points form an ellipse with focal points at the source and receiver positions.

If we look carefully at the gradient expression.

$$g = \int_{t=0}^{t=T} \sum_{s=1}^{s=N_s} \sum_{r=1}^{s=N_r} \left(P \frac{\partial u(x,t)}{\partial v} \right)_r^{\dagger} (Pu(x,t) - d_{obs}(x,t)) \tag{3.101}$$

The u is the wavefield propagating in the medium, generated by source s. The diffracted wavefield has as a virtual source the term $-(\frac{\partial A}{\partial m})u$. The term $\left(P \frac{\partial u(x,t)}{\partial v} \right)_r^{\dagger}$ is diffracted at one receiver position r.The gradient g is the multiplication of the diffracted wavefield with the residual data $(Pu - d)$. (Figure 3.10).

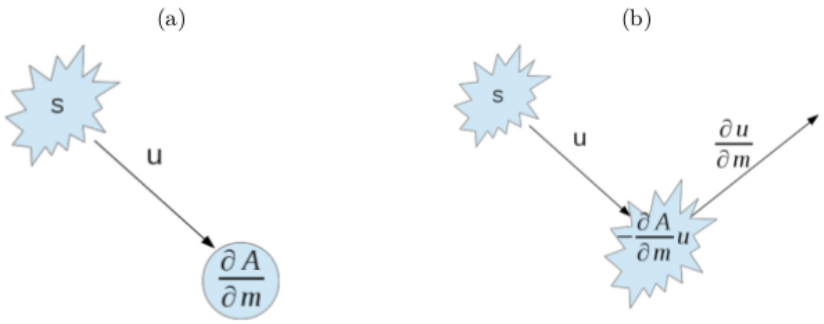


Figure 3.10: a) $Au = s$, source s generates a wavefield that is propagated by operator A b) $A\frac{\partial u}{\partial v} = -\frac{\partial A}{\partial v}u$, source $-\frac{\partial A}{\partial v}u$ generates diffracted wavefield $\frac{\partial u}{\partial v}$ propagated by operator A

The i_{th} component of the gradient will be non-zero if the diffracted wavefield by hetereogenity v_i arrives at the receiver position at the same time at as the recorded time for the residual. Consider a true velocity model that consists of an homogeneous background v_0 and three scattering points. There is one source denoted by a star, and a line of receivers at the surface at the positions of the dashed line. The source s generates a wavefield that travels through the model v and that will be diffracted by each of the heterogeneities. In figure the partial derivative wavefield $\frac{\partial u}{\partial v}$ (diffracted by the green scatterer) at the receiver positions is plotted. In , the data residuals recorded at the receiver positions are shown. Three arrivals are seen due to the three diffraction points. Recall that the gradient of equation is the correlation $\frac{\partial u}{\partial v}$ and the residuals. The sum over time of the correlation the partial derivative wavefields and the data residuals is going to generate a non-zero gradient value at point v_i , because we can see that the two arrivals coincide perfectly.(Figure 3.11)

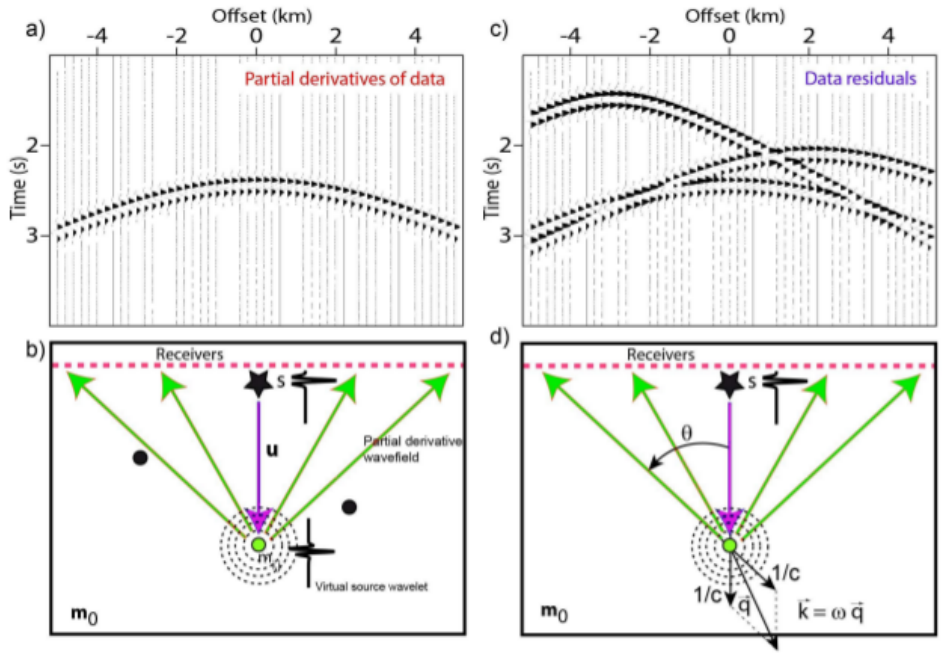


Figure 3.11: Figure taken from (Operto et al., 2013) [2]. The true velocity model consists of an homogeneous background m_0 and three scattering points. There is one source denoted by a star, and receivers at the positions of the dashed line. b) The source s generates a wavefield that travels through the model m and is going to be diffracted by each of the heterogeneities. a) The partial derivative wavefield $\frac{\partial u}{\partial m}$ (diffracted by the green scatterer) at the receiver positions. c) The data residuals recorded at the receiver positions. Three arrivals are seen due to the three diffraction points. The sum over time of the correlation the partial derivative wavefields (figure a) and the data residuals (figure c) is going to generate a non-zero gradient value at point v_i .d) The role of the acquisition can be analyzed, considering the angle between the source and the receiver.

3.0.6.2 The Hessian: Computaion and Interpretation

The Newton descent direction methods require the computation of Hessian , second derivative of the misfit function χ

$$H = \nabla_m^2 \chi = \left(P\frac{\partial u}{\partial m}\right)^\dagger \left(P\frac{\partial u}{\partial m}\right) + \left(\frac{\partial u}{\partial m}\right)^\dagger P^\dagger (Pu - d) \tag{3.102}$$

It involves the computation of double diffracted wavefiedls and it is quite expensive to calculate . It can be calculated either differentiating forward problem twice

$$A\frac{\partial^2 u}{\partial m_i \partial m_j} = -\frac{\partial A}{\partial m_i}\frac{\partial u}{\partial m_j} - \frac{\partial^2 A}{\partial m_i \partial m_j}u - \frac{\partial A}{\partial m_j}\frac{\partial u}{\partial m_i} \tag{3.103}$$

or using adjoint state method . However we will not discuss here as we didn't employ Newton methods but for the completion we will discuss the physical meaning of the Hessian term.

3.0.6.2.1 Physical Interpretation :The Hessian allows to correct artefacts in the imaging condition. The artefacts in the gradient will come mainly from the limited bandwidth of the source, the smearing artefacts of the imaging condition along an ellipse and double scattered energy that is mapped as single scattered energy. The first term in the Hessian represents the zero lag cross correlations of the partial derivative wavefields (diffracted at different points in the model) at the receiver positions. Let N be a neighborhood containing all the model parameters around $m_i j$. The size of this neighborhood will depend on the source spectrum. For all the points in this neighborhood, all the partial derivative wavefields $\frac{\partial u}{\partial m}$ will arrive at similar times to the receiver positions and thus all will correlate perfectly with the residuals . This will cause a defocusing, because all the points around m i will be imaged, giving rise to rather smooth models. As explained by Pratt et al. (1998)[3], the structure of the GN approximation of the Hessian is similar to a convolutional or smoothing operator. The application of the inverse Hessian sharp- ens and refocuses the gradient, just as a spiking deconvolution in seismic data processing.
Even in the case of a perfect source with δ as a spectrum, there would still be artefacts in the gradient due to incomplete illumination. For one source-receiver pair, all the points along the ellipse that have the source and receiver as a foci, will be imaged. If there are enough sources and receivers in an adequate acquisition geometry, destructive interference will cancel all points except the target we wish to image. However, with a limited acquisition these artefacts may persist. The Figure 3.12 shows an example done by Pratt

et al. (1998)[3] , where the gradient has the smile artefacts present. When the inverse GN approximation of the Hessian is applied, these artefacts disappear.

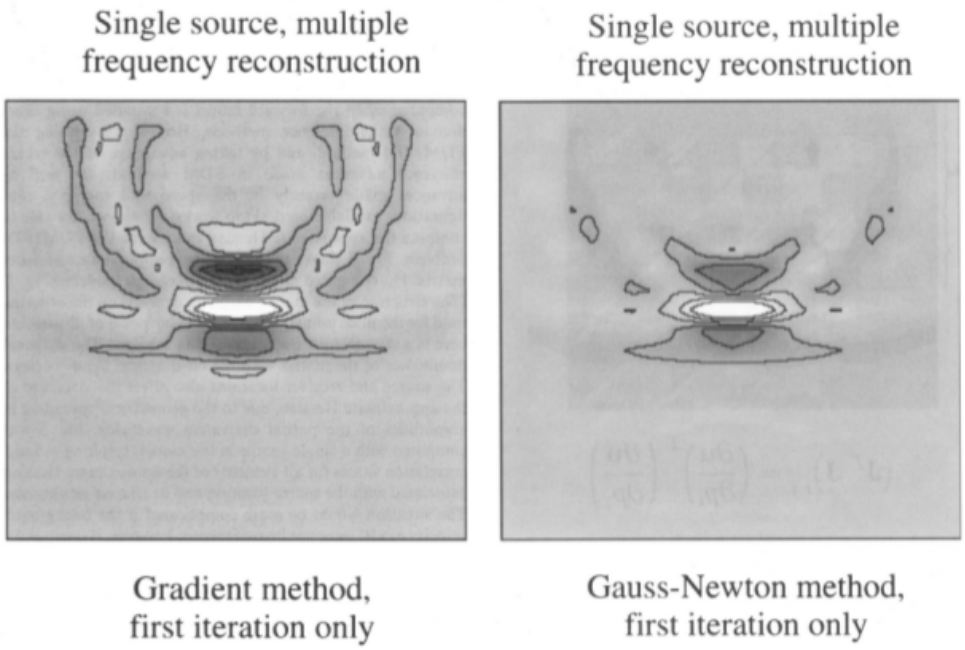


Figure 3.12: Figure from Pratt et al. (1998)[3]. With one source and receivers on the surface. On the left, gradient in the first iteration shows the smile artefacts. On the right, the Gauss-Newton descent direction has removed the smile artefacts.

The second term in the Hessian in equation is less important if either the residuals are small, or the double diffracted wavefields are small (meaning we are in the linear regime). That is, the second order partial derivative wavefields are important when a change in a model parameter can affect the partial derivative wavefields.

3.0.6.3 Discussion on Image Resolution

The resolution analyses is directly related to the imaging condition and the Fresnel zone concepts.However it could be better understood in frequency - wavenumber domain.
In the frequency domain, the imaging kernel will show the ellipses that satisfy the imaging condition for all the arrivals in time. An example of an imaging kernel for one frequency (sometimes known as sensi-tivity kernel) is represented in Figure . The dark and light

fringes denote the sensitivity of the model parameters to one frequency. The central ellipse defines the first Fresnel zone. The first Fresnel zone is the region around the direct path such that the travel time for points in this region differs in no more than $T/2$. All arrivals of scattered paths within half a period of one arrival will generate constructive interference in the seismograms. In terms of wavelength, the Fresnel zone is the region around the reflector path such that the ray path differs by no more than $\lambda/2$. The first Fresnel zone defined for each offset $D$is important for two reasons. If the background model is not sufficiently accurate such that the difference in between the arrival of scattered modeled data and the recorded data does not exceeds more than half a period, the resulting data residual will be back-propagated on the wrong Fresnel zone leading to erroneous update of the subsurface model. This is referred to as cycle-skipping. The width of the first Fresnel zone is $\sqrt{\lambda D}$ (Williamson, 1990)[32].

The second reason why the Fresnel zone is important is because it limits the resolution of the final image. Two point scatterers that generate reflected waves whose arrival times is less than half a period, will not be distinguishable. For a surface acquisition, the smallest distance dx and dz between two point scatters for them to have an arrival time difference of at least $T/2$ is approximately (Schuster , 2007) [33].

$$dx \approx \lambda\sqrt{D^2 + z^2/(4D)} \tag{3.104}$$

$$dz \approx \lambda/4 \tag{3.105}$$

where z is the depth of the scatterer and D is the offset. From these equations we can see that the vertical resolution dz of the image does not depend on the offset, but only on the smallest wavelength. The horizontal resolution dx is different for shallow or deep scatterers. For example, for deep scatterers ($z >> D$), long offsets D are needed to improve the horizontal resolution.

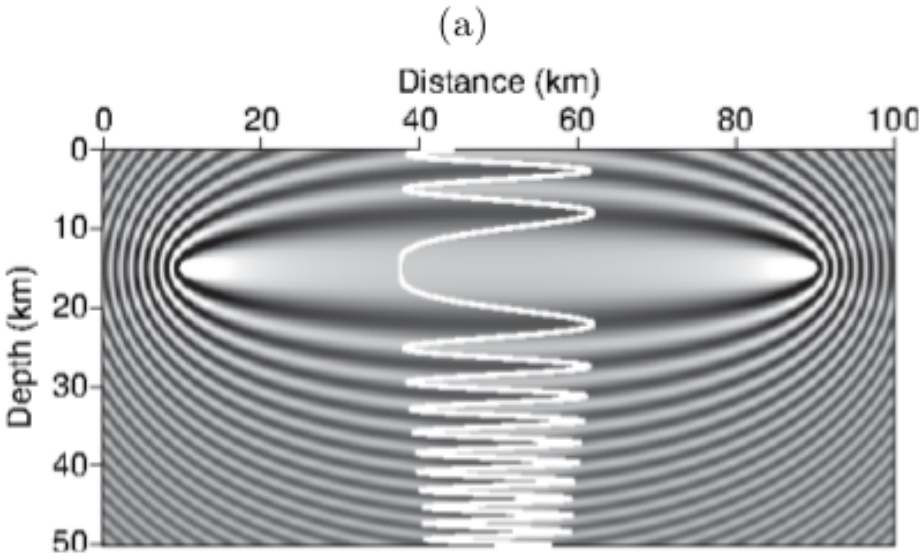


Figure 3.13: a) Imaging kernel in the frequency domain for one source receive pair. The first central ellipse denotes the first Fresnel zone. The black and white fringes describe the iso-phase surfaces of all the scattered arrivals. (Each isochrone represents all the parameters in the model space that can explain a seismic phase arrival at time t). The width of the isochrones decreases in depth as the scattering angle decreases (Figure from : Woodward, 1992)[4]

3.0.6.3.1 Wavenumber Reconstruction Another aspect of the quality of the image concerns the capability to reconstruct small and large details of the model. By applying a Fourier transform to the depth and horizontal distance variables (z, x), and working in the wavenumber domain (k_z, k_x) it is possible to analyze the resolution of the model in terms of coverage of the wavenumber spectrum.To describe the transmission components over a wide range of offsets the low-wavenumber components of the velocity model must be accurate. To account for reflection amplitudes, the high wavenumber components of the velocity model must be correct (Mora, 1989)[7].

Considering an inversion in the frequency domain (under far field assumptions and using the Born approximation), and using a delta a source function, the direct $u(x, \omega)$ and

back-propagated wavefields $\lambda(x,\omega)$ in terms of the Green's functions are

$$u(x,\omega) = G(x,s) \tag{3.106}$$

$$\lambda(x,\omega) = G^{\dagger}(x,r)(Pu-d) \tag{3.107}$$

plugging back into the gradient expression we get

$$\nabla_m \chi = \omega^2 \sum_{n=1}^{n=N_s} \sum_{n=1}^{n=N_r} (G(x,s)G^{\dagger}(x,r)(Pu-d)) \tag{3.108}$$

Under plane wave approximation

$$\nabla_m \chi = \omega^2 \sum_{n=1}^{n=N_s} \sum_{n=1}^{n=N_r} (\exp(-i\vec{k}(\hat{s}+\hat{r}))(Pu-d) \tag{3.109}$$

the unit vector $\hat{s}$ points from the source to the diffracting point x and $\hat{r}$ is a unit vector from the receiver to the diffracting point.The gradient can be interpreted as an inverse Fourier where the coefficients are given by the residuals. Let us denote $\vec{k_{S,x}} = \vec{k}.\hat{s}$ as the vector from the source to point x, and $\vec{k_{R,x}} = \vec{k}.\hat{r}$ as the vector from the receiver to the point x. In Figure 3.14, the wave vectors are plotted, in the wave vector space. The angle between $\vec{k_{S,x}}$ and $\vec{k_{R,x}}$ at the diffracting point x is θ. The resulting wave vector at x is

$$\vec{k} = \frac{\omega}{|v|}\cos(\theta/2)\hat{n} \tag{3.110}$$

$$\hat{n} = \frac{\vec{k_{S,x}} + \vec{k_{R,x}}}{(\sqrt{||\vec{k_{S,x}} + \vec{k_{R,x}}||^2}} \tag{3.111}$$

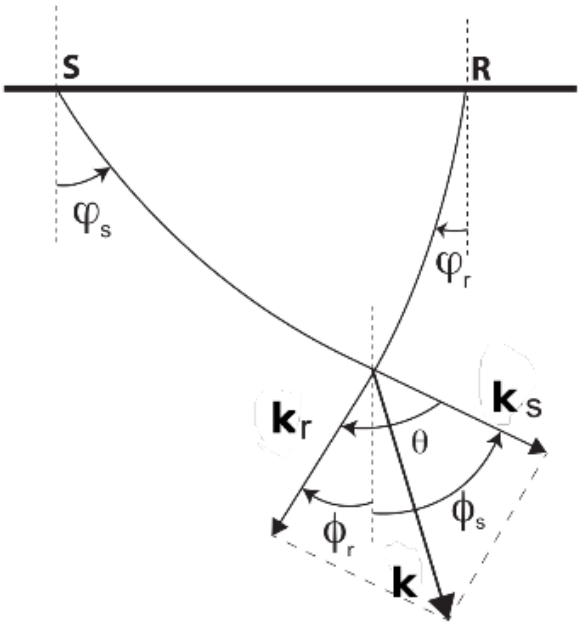


Figure 3.14: Figure adapted from Sirgue and Pratt (2004)[5]. $k_{S,x}$ and $k_{R,x}$ are vectors from the source to the diffracting point and from the receiver to the diffracting point.

In order to have a gradient that can reconstruct all wave numbers both in amplitudes and directions, different angles and different frequencies must be considered. The term ω and the magnitude of v play a role in the magnitude of the reconstructed wave number, and the angle θ plays a role in the direction and magnitude.

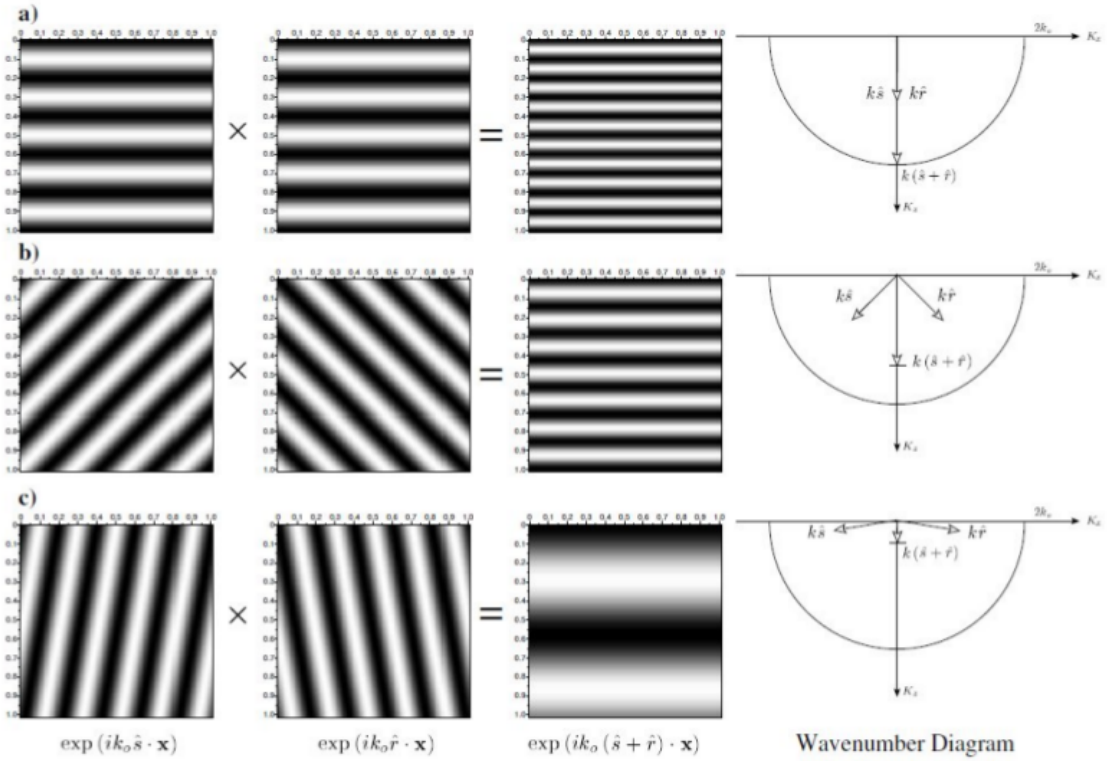


Figure 3.15: Figure from Sirgue (2003)[6]. Under the plane wave approximation, an illustration of the variation of the wave number contribution in the gradient, for different incidence-scattering angles. The first column represents the wave paths from the source to the scattering point $\exp(i\vec{k}.\hat{s})$. The second column represents the wave paths from the receiver to the scattering point $\exp(i\vec{k}.\hat{r})$. The last column is the multiplication of both paths $\exp(i\vec{k}.(\hat{s}+\hat{r})$, which is proportional to the gradient. a) The source are receiver are at the same position. The incidence and scattering angle is 0^0. The total wave number is high, which can be seen in the high resolution of the gradient. b) The incidence and scattering angle is 90^0 . c) The incidence and scattering angle is 160^0 . The lowest wave number contribution is given the high incidence and scattering angles. The high wave number contribution is given by small incidence and scattering angles

In Figure 3.15 we can see the effect of changing the angle θ on the wave number reconstruction, for a fixed frequency. The angle is increased from Figure to Figure 3.15c, and as a consequence the reconstructed wave number decreases.

The wavenumber resolution capacity k_x, k_z for a given source receiver-pair therefore depends on the imaging condition, the acquisition and position of the reflector (that determine θ), the magnitude of the propagation velocity of the waves v and the frequency f . A representation of the usual wave number illumination is shown in 3.16 , using infinite offset data. There is a gap in low horizontal and vertical number spectrum. For example, note that for $k_z = 0$, there is no resolution in k_x . On the other hand, if $k_x = 0$, the whole k_z can be recovered. This is a consequence of the fact that we are using a surface acquisition and so, the resultant sum of the wavevector usually has k_z as its most important component.

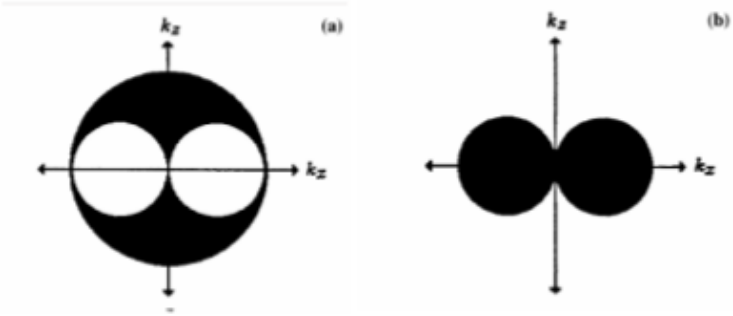


Figure 3.16: Figure from Mora (1989)[7]. a) Shows the part of the usual part of wave number spectrum in the image space that can be resolved with the infinite offset data. As it is clear, there is a gap in low number spectrum. b) Shows the part of the wave number spectrum in the image that can be recovered by introducing reflectivity information in the model, and expanding the imaging condition , in forward and back-scattered direct and back-propagated wavefields.

To try to recover the low wavenumber (tomographic information), Mora (1989)[7] suggested to introduce reflectivity information in the model (Mora, 1989)[7]. The idea is that the reflector will create back-scattered energy. Therefore, the waves will not only coincide in a point x, as shown in Figure , but the reflected direct wave will meet the

back propagated wavefield from the receiver at other points, as illustrated in Figure . This configuration resembles a tomographic configuration where the reflector would correspond to the source and the receivers are on the opposite side. The resultant wavenumber has a low magnitude $|\vec{k}|$. By doing this, it is possible to recover the low wavenumber spectrum and fill the gap as shown in (Mora, 1989)[7]. In general, in the presence of strong reflectors, the direct and back-propagated wavefields can be separated in those travelling downward and upward, $u = u^+u^-$ and $\lambda = \lambda^+\lambda^-$, where $+$ indicates downward and $-$ indicates upward travelling. The correlation of u and λ will therefore provide four terms. The correlation of wavefields travelling in opposite directions will cover the low part of the wavenumber spectrum.

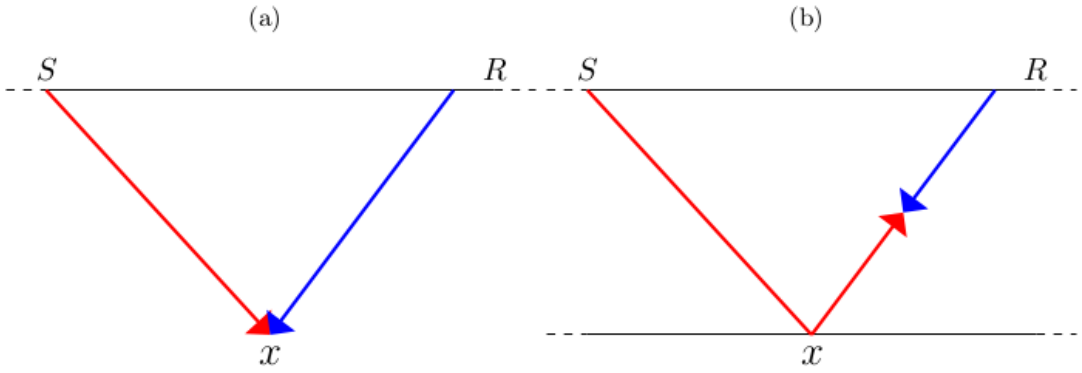


Figure 3.17: Imaging condition with and without a reflector in the background model. a) correlation of u and λ. Reconstructs high magnitude wavenumbers. b) Correlation of back-scattered u and λ. Reconstructs low magnitude wavenumbers.

3.0.7 CUDA and Parallelisation strategies in FWI

3.0.7.1 Introduction

There is a need for parallelization in the domain of scientific computing. Microprocessors based on a single central processing unit (CPU), such as those in the Intel Pentium family brought GFLOPS, or giga 10^{12} floating-point operations per second, to the desktop

and TFLOPS, or tera (10^{15}) floating-point operations per second, to cluster servers. The users, in turn, demand even more improvements once they become accustomed to these improvements, creating a positive (virtuous) cycle for the computer industry. This drive, however, has slowed since 2003 due to energy consumption and heat dissipation issues that limited the increase of the clock frequency and the level of productive activities that can be performed in each clock period within a single CPU. Since then, virtually all microprocessor vendors have switched to models where multiple processing units, referred to as processor cores, are used in each chip to increase the processing power. This switch has exerted a tremendous impact on the software developer community.

Many-threads processors, especially the GPUs, have led the race of floating-point performance since 2003. As of now the ratio of peak floating-point calculation throughput between many-thread GPUs and multicore CPUs is more than 10. These are not necessarily application speeds, but are merely the raw speed that the execution resources can potentially support in these chips: 1.5 teraflops versus 150 gigaflops double precision in 2012. To date, this large performance gap has already motivated many applica- tion developers to move the computationally intensive parts of their software to GPUs for execution. Not surprisingly, these computationally intensive parts are also the prime target of parallel programming—when there is more work to do, there is more opportunity to divide the work among cooperating parallel workers.

GPU's have different design philosophy than CPU design. CPU is optimized for sequential code performance . It makes use of sophisticated control logic to allow instructions from a single thread to execute in parallel or even out of their sequential order while maintaining the appearance of sequential execution. More importantly, large cache memories are provided to reduce the instruction and data access latencies of large complex applications. Neither control logic nor cache memories contribute to the peak calculation speed. As of 2012, the high-end general-purpose multicore microprocessors typically have six to eight large processor cores and multiple megabytes of on-chip cache memories designed

to deliver strong sequential code performance.

Memory bandwidth is another important issue. The speed of many applications is limited by the rate at which data can be delivered from the memory system into the processors. Graphics chips have been operating at approximately six times the memory bandwidth of contemporaneously available CPU chips.

The design philosophy of GPUs is shaped by the fast-growing video game industry that exerts tremendous economic pressure for the ability to perform a massive number of floating-point calculations per video frame in advanced games.The prevailing solution is to optimize for the execution throughput of massive numbers of threads. The design saves chip area and power by allowing pipelined memory channels and arithmetic operations to have long latency. The reduced area and power of the memory access hardware and arithmetic units allows the designers to have more of them on a chip and thus increase the total execution throughput.

FIGURE 1.1

CPUs and GPUs have fundamentally different design philosophies.

Figure 3.18: CPU vs GPU design (source:[8])

3.0.7.1.1 Programming Model CUDA is a high level language almost like C but with a few straight forward extensions. The extensions to the C programming language are four-fold:

a) Function type qualifiers to specify whether a function executes on the host or on the de-

vice and whether it is callable from the host or from the device ___*global*___, ___*device*___, ___*host*___.

b) Variable type qualifiers to specify the memory location of a variable___*device*___, ___*shared*___.

c) A new directive to specify how a kernel is executed on the device from the host. Four built-in variables that specify the grid and block dimensions and the block and thread indices - gridDim, blockIdx, blockDim, threadIdx.

CUDA encapsulates hardware model, so we don't have to worry about hardware model changes, all the conveniences of C vs assembly. Learning the hardware and developing parallel algorithms is still difficult. But the infrastructure for writing, developing,debugging and maintaining source code is straight forward and similar to conventional serial programming. Steps in a CUDA code :

1) Initialize/acquire device (GPU)

2) Allocate memory on GPU

3) Copy data from host to GPU

4) Execute kernel on GPU

5) Copy data from GPU to host

6) Deallocate memory on GPU

7) Release device

A kernel is a function callable from the host and executed on the CUDA device – simultaneously by many threads in parallel. How to call a kernel involves specifying the name of the kernel plus an execution configuration. An execution configuration just means defining the number of parallel threads in a group and the number of groups to use when running the kernel for the CUDA device

3.0.7.1.2 Memory Model Each CUDA device has several memories that can be used by programmers to achieve high Computation to Global Memory Access (CGMA) ratio and thus high execution speed in their kernels. Variables that reside in registers and shared memories can be accessed at very high speed in a highly parallel manner.

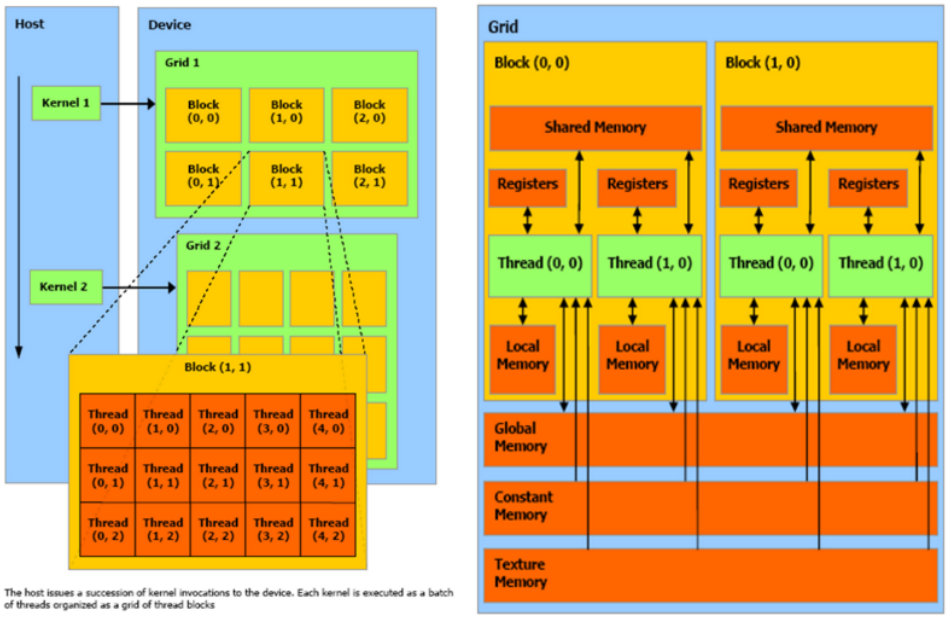


Figure 3.19: CUDA Memory model (source:[8])

Registers are allocated to individual threads; each thread can only access its own registers. A kernel function typically uses registers to hold frequently accessed variables that are private to each thread. Shared memories are allocated to thread blocks; all threads in a block can access variables in the shared memory locations allocated to the block. Shared memories are efficient means for threads to cooperate by sharing the results of their work. At the middle of the table, we see global memory constant memory. These are the memories that the host code can write (W) and read (R) by calling API functions. The global memory can be accessed by all the threads at anytime of program execuution. The constant memory allows read-only access by the device and provides faster and more parallel data access paths for CUDA kernel execution than the global memory.

Memory	Location	Cached	Access	Who
Local	Off-chip	No	Read/Write	One Tthread
Shared	On-chip	N/A	Read/write	All threads in a block
Global	Off-chip	No	Read/write	All threads + CPU
Constant	Off-chip	Yes	Read	All threads + CPU
Texture	Off-chip	Yes	Read	All threads + CPU

Figure 3.20: CUDA Memory types (source:[8])

3.0.7.2 Parallel Strategies in FWI

In Full waveform inversion , the storage of forward wavefields is huge problem . The traditional way is to record the forward wavefields to the disk so that it could be read again while applying the imaging condition. It requires frequent $I\emptyset$ in our algorithm .Despite of high memory band width in CUDA it is still costly , so we will utilise wavefield reconstruction method . The wavefield reconstruction method is a way to recover the wavefield via backward reconstructing or forward remodeling, using the saved wavefield shots and boundaries. It is of special value for GPU computing because saving the data in device variables eliminates data transfer between CPU and GPU. The other strategies we will discuss is the use of shared memory for forward modeling and an efficient parallell reduction for calculating objective function and step size in the optimization problem

3.0.7.2.1 Forward Modeling in CUDA I employed a tiling approach in which each thread block loads a tile of data from the multidimensional grid into shared memory, so that each thread in the block can access all elements of the shared memory tile as needed. The size of the tiles must be chosen taking into account ,the warp size and the amount of shared memory available on the device (should be a multiple of 16, half-warp size). The finite-difference coefficients for higher order approximation can be declared in constant memory if necessary since as they are constants and they will be broadcasted to each warps.

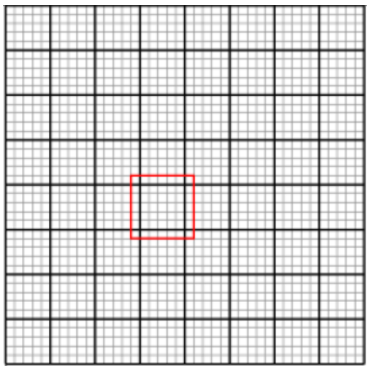


Figure 3.21: Red block size shows the shared memory

In the forward modeling kernel ,the pressure wavefields corresponding to the two time steps are loaded into share memory in the form of tiles . The boundary conditions used are Clayton and Enquist A2 boundary conditions.The Tile Size in both x and z direction is 16 .

3.0.7.2.2 Wavefield Reconstructed Method We used wavefield reconstruction method to reconstruct the source wavefields in the backpropagation step while applying the imaging condition. It is based on the intution that if forward modeling can be written as

$$p^{k+1} = 2p^k - p^{k-1} + v^2 \Delta t^2 \nabla^2 p^k \tag{3.112}$$

then wavefields could be reconstructed by just swapping the terms p^{k+1} and p^{k-1}

$$p^{k-1} = 2p^k - p^{k+1} + v^2 \Delta t^2 \nabla^2 p^k \tag{3.113}$$

The wavefield can be extrapolated by just swapping the pointers .

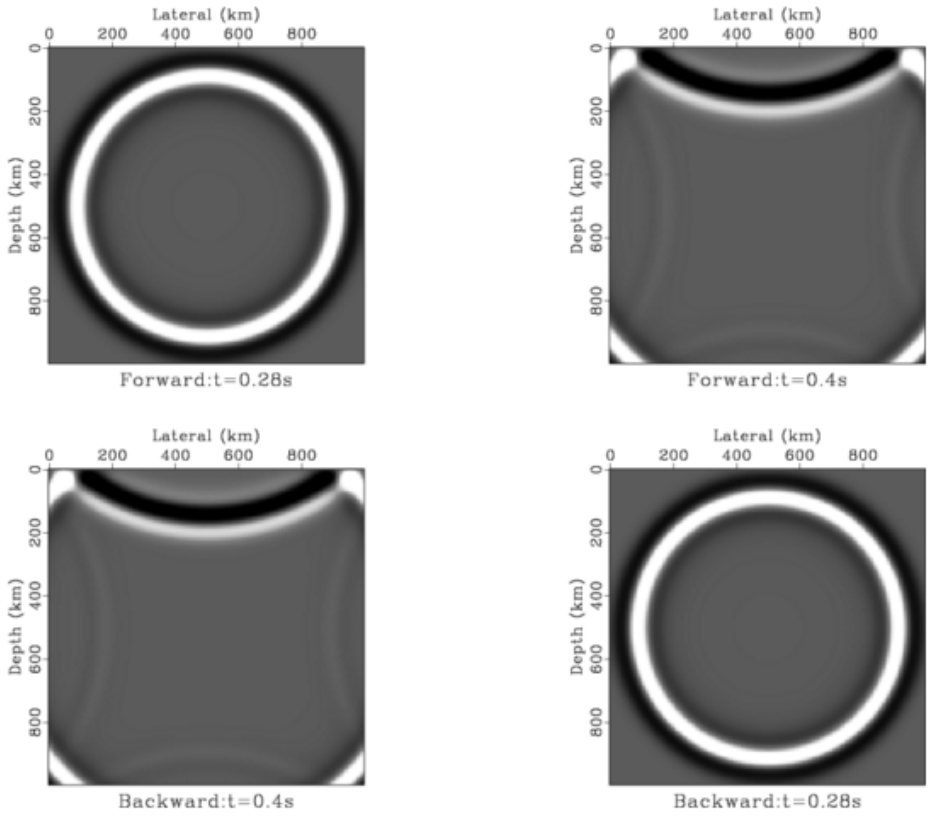


Figure 3.22: Wavefield Reconstruction

In figure , we tested the results for model size 200 by 200 with dx=5 and dz=5 . A ricker source with peak frequency is placed at the center of the model. We used free surface conditions for this test case and we record wavefields at t=0.28 s and t=0.4 s. As you can see that the wavefields are exaclty reconstructed.

3.0.7.2.3 Parallel Reduction : The parallel reduction could be used in calculating objective function , finding the step size and descent direction in FWI algorithm . However there are certain things which should be taken into conideration while implementing the reduction on CUDA. The programmer should be familiar with CUDA performance characteristics which are :

1) **Memory Coalescing**

2) **Divergent Branching**

3) **Memory Bank Conflicts**

4) **Latency hiding**

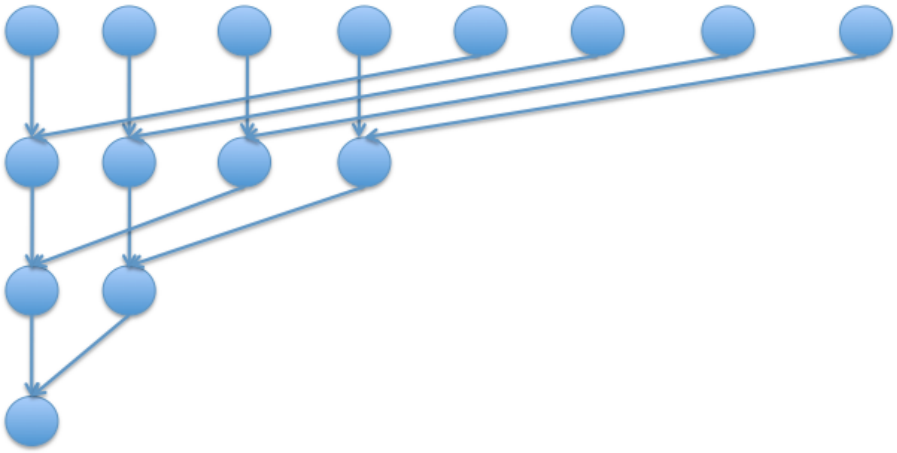


Figure 3.23: Parallel Reduction

A sequential addressing scheme is utilized because it is free of conflict (Harris et al., 2007). The parallel reduction approach builds a summation tree to do stepwise partial sums. In each level half of the threads will perform reading from global memory and writing to shared memory. The required number of threads will decrease to be half of previous level. We unrolled the last warps to further increase the efficiency.

Chapter 4

Numerical results : FWI

The code developed for Full waveform Inversion is tested on Marmousi model . The source used is Rikcer source with peak frequency 10 Hz . The algorithm used for inversion is conjugate gradient method.

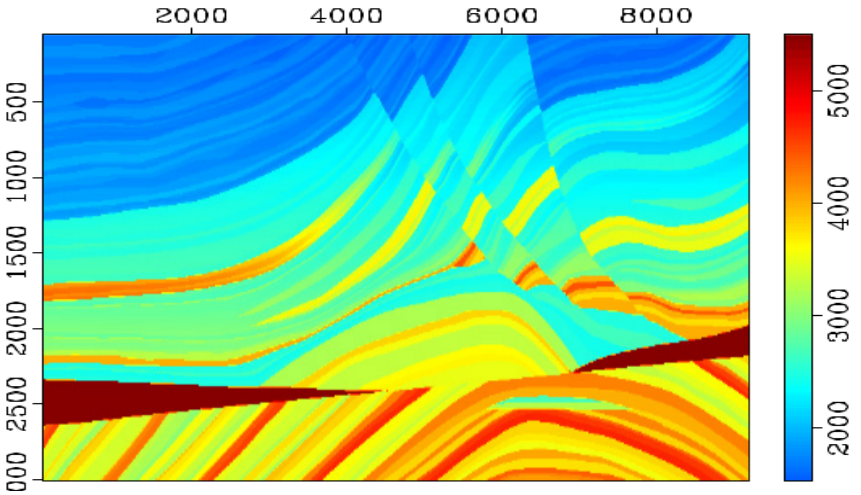


Figure 4.1: Marmousi Velocity Model, (IFP)

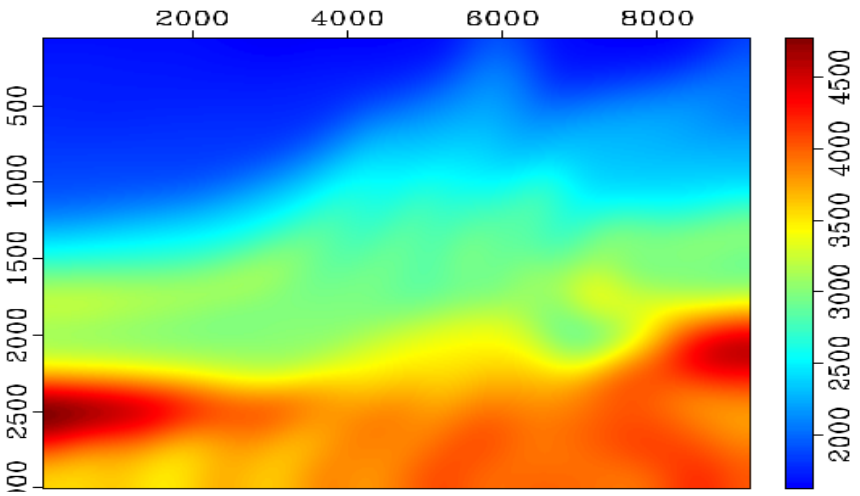


Figure 4.2: Initial Velocity Model

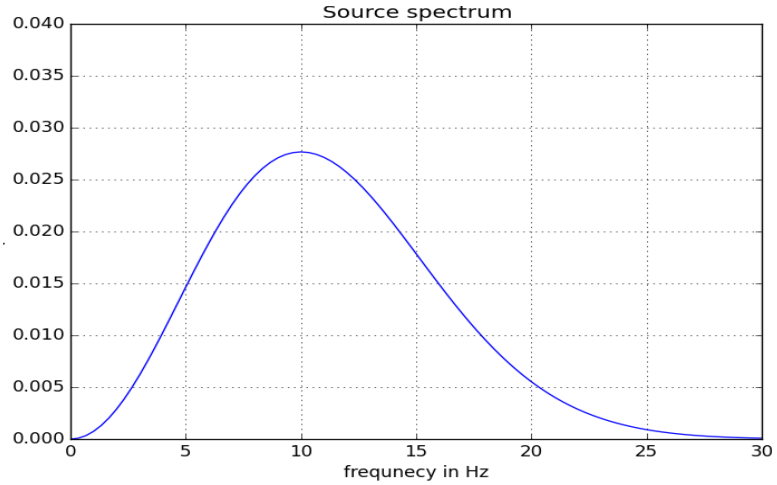


Figure 4.3: Frequencies involved in inversion

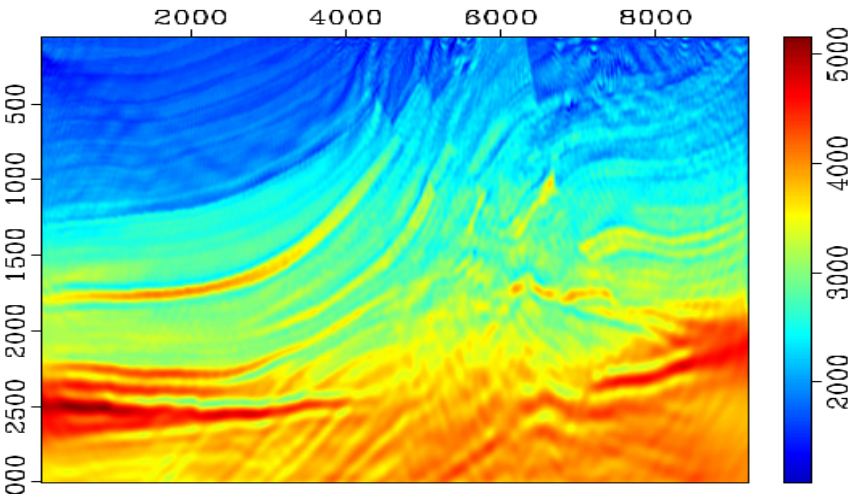


Figure 4.4: Velocity Model recovered after 100^{th} iteration

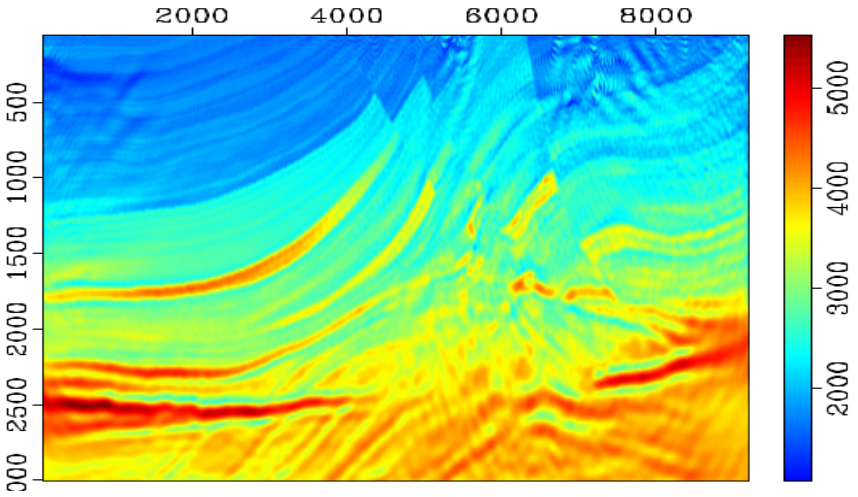


Figure 4.5: Velocity Model recovered after 250^{th} iteration

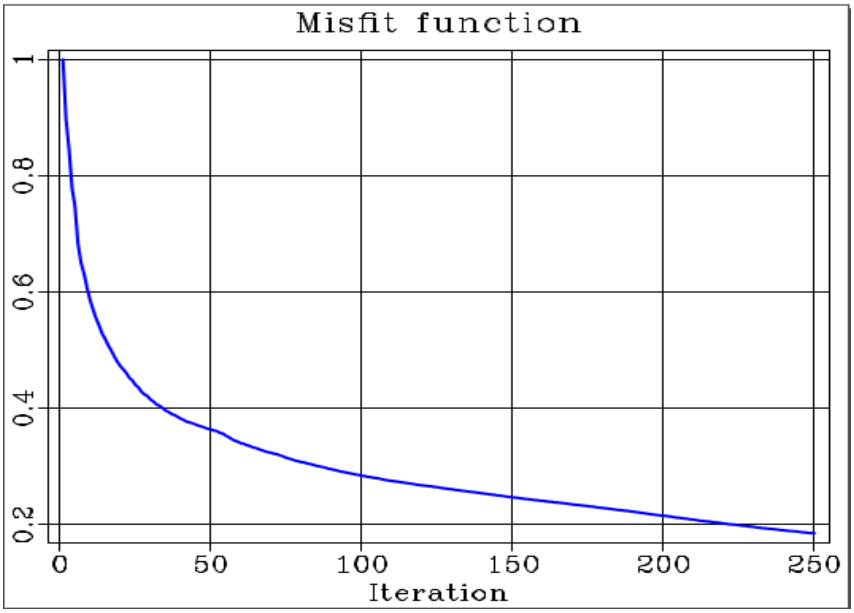


Figure 4.6: Normalised Objective function value

Modeling Parameters :

Model size: 251×767

No of receivers : 767

No of sources : 60

No of iterations: 250

Time per iteration : 60 sec

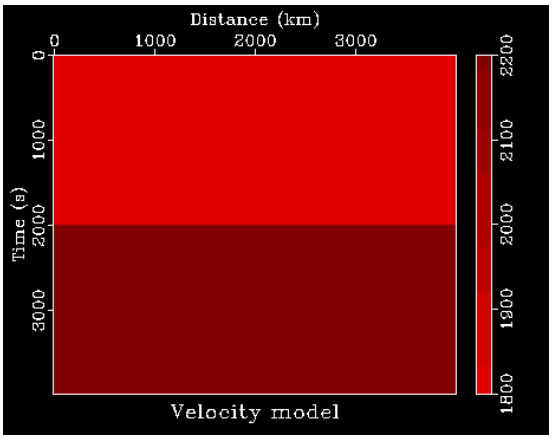
Chapter 5

Specularity + FWI

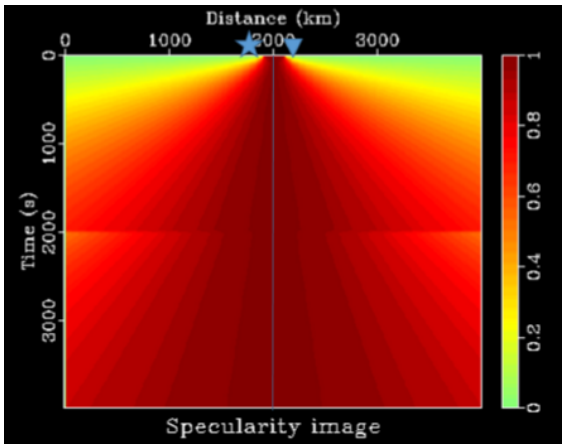
Specularity is defined as a measure of deviation from Snell's law at the image point. Moser and Howard (2008) defined specularity S as the dot product of the angle bisector of incident and reflected ray with the normal vector to the surface.

$$S(s,r,x) = \frac{(P_s + P_r).n}{|P_s + P_r|} \tag{5.1}$$

where P_s and P_r are the sourece side and receiver side slowness vectors , obtained from the travel time derivatives , n is the normal unit vector to the reflector which can be calculated from conventional depth migrated image. It was defined as a weighting function ,in terms of single source-receiver pair ,to suppress the contribution from inside the Fresnel zone , in Kirchhoff based integrals such that it only images the nonspecular events corresponding to diffractions thus removing the contribution from reflections. For specular reflection S should be equal to 1 and for non-specular events $S < 1$ which is part of our interest. We will extend the definition to the full acquistion system and also attempt extracting source and receiver side slowness vectors using poynting vectors for the wavefield definition.
Let's say we have a layered velocity model with single reflector and single source-receiver pair at the surface.

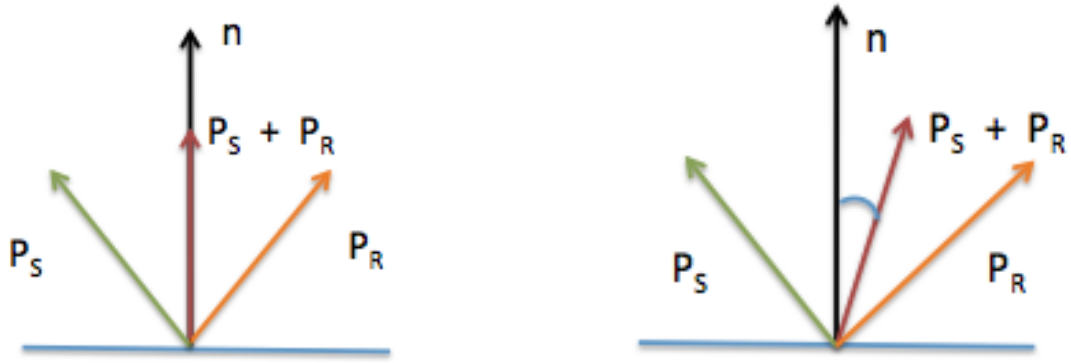


(a) layered velocity model



(b) specularity image

For a single reflector case as one can see that specularity decreases away from the points along the mid-point of the source-receiver pair which one would expect because snell's law only holds perfectly for the points that falls along the vertical straight line passing through the mid-point . Since the fresnel zone will fall over a certain area rather than a single point so the points in the vicinity of this middle point (at the true event location) will be treated as part of specular reflections or the criteria for specularity will be a bit relaxed. The points that fall's outside this region will behave as potential diffractors for a single-receiver case , even if they fall's on the reflector as there no receiver's to record their specular behaviour. The specularity in the above example is computed using true velocity model dip field just for the illustration purposes however it can be calculated using the conventional depth image.



(a) specular reflection

(b) non-specular diffraction

5.0.8 Ray based Specularity

In this section we will try to extend the definiton for full acquistion system to design a specularity based filter which can be applied for cleaning the gradient in Full waveform inversion or in Reverse Time migrated image to clean the specular part. Here we will demonstrate using Reverse time migrated images .

Let us review the example of single reflector which we used in the previous section . In this case we will backpropagate rays simultaneously from receivers and use single source ray field.

$$S^T(s,r,x) = \frac{(P_s + P_r).n^T}{|P_s + P_r|} \tag{5.2}$$

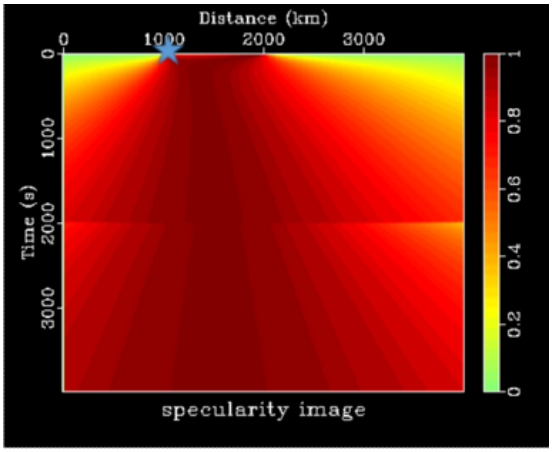
where P_s and P_r are the source side ray field and receiver side ray field respectively.

P_s and P_r can be calculated by taking derivative of traveltimes from source side or receiver side to each poin in the domain . Travel times can computed by solving the eikonal equation.

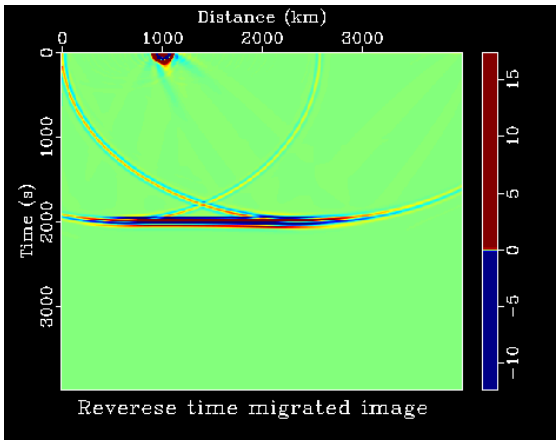
$$(\nabla_x T)^2 + (\nabla_z T)^2 = \frac{1}{v^2} \tag{5.3}$$

The dip field can be calculated using plane wave destruction filters or local slant stack method and using the simple trignometric identites , normal vector components could be

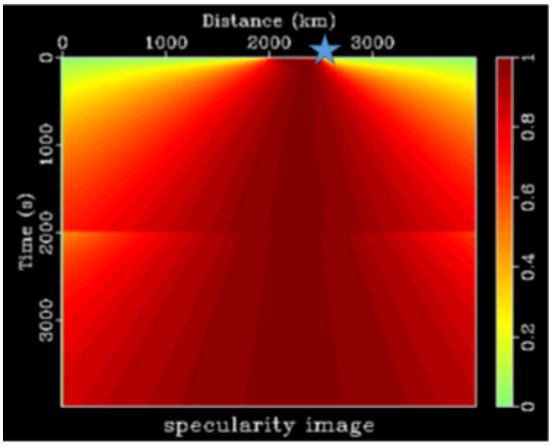
computed.



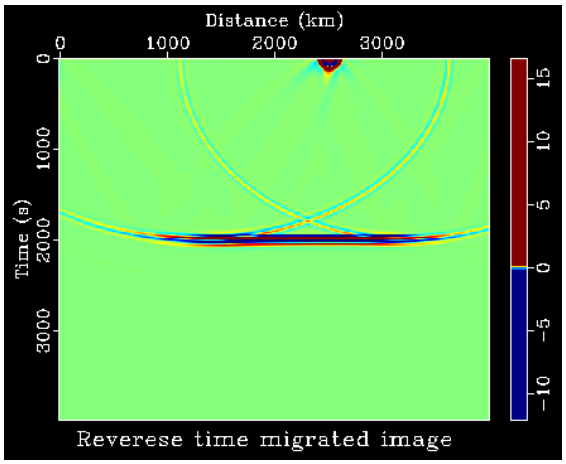
(a) specularity image



(b) RTM contribution due to same source



(a) specularity image



(b) RTM contribution due to same source

Specularity is basically an anti-stationary filter which suppress the contributions from stationary specular part.It works in the same way as it did in Kirchhoff sense and is also limited by finite frequencies and finite-aperture acquisition.For dip calculation and travel-times programs from Madagascar were utilised which are freely available on the Internet. The tapering value for specularity cleaning was decided using the specularity image and ranges from 0.6 to 0.8 in our test examples.

Example-1 : Step Layer Model

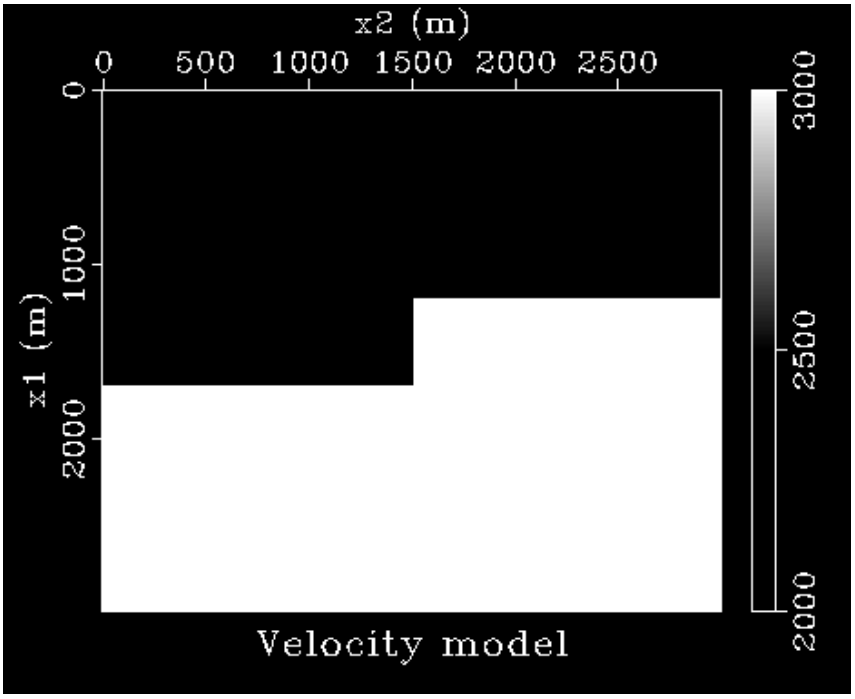


Figure 5.5: Velocity model

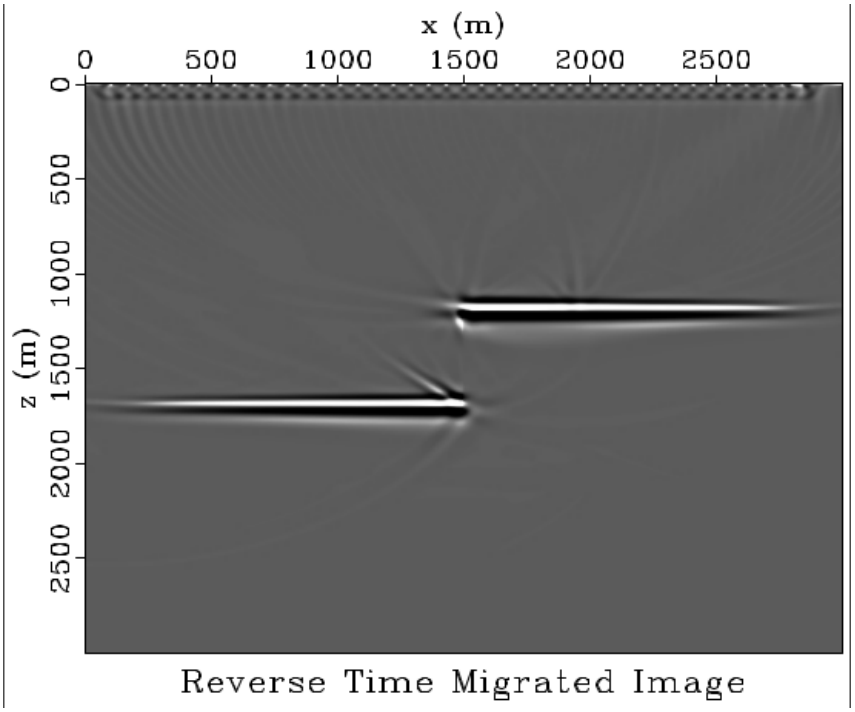


Figure 5.6: RTM image

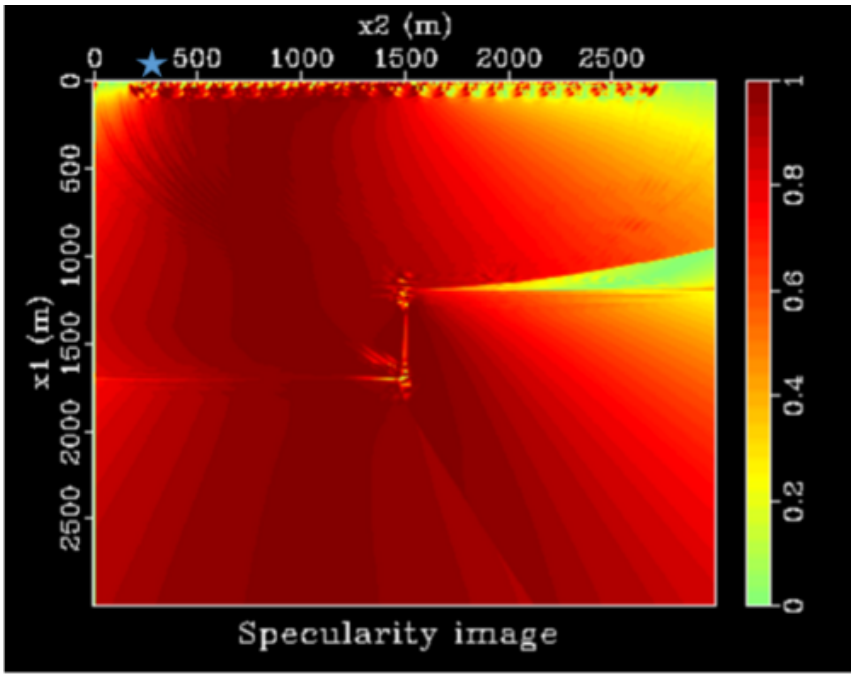


Figure 5.7: Specularity image

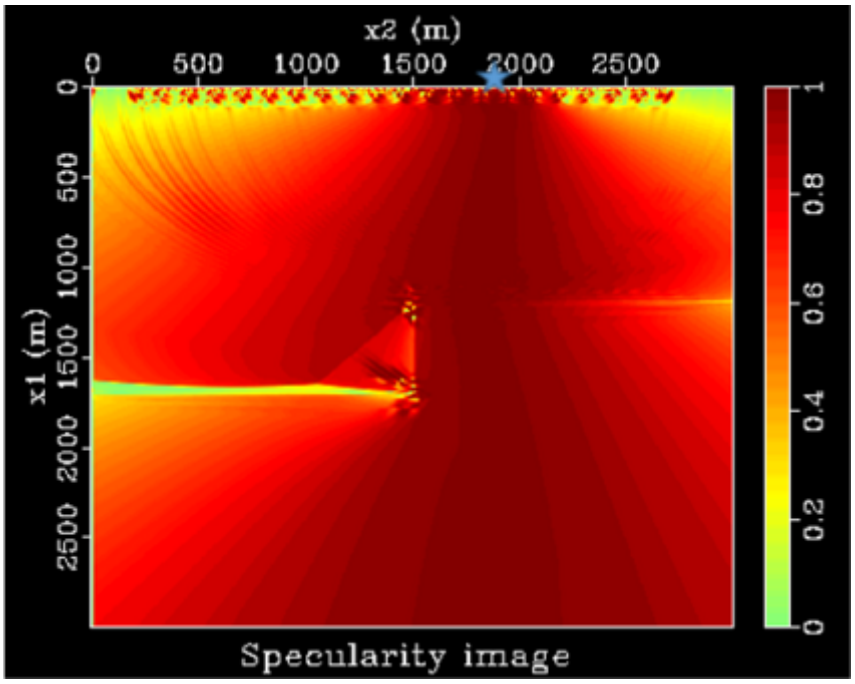


Figure 5.8: Specularity image

Figure 3 an 4 is specularity calculated over whole model domain due to a single source (position denoted by star) and backpropagated ray field. The low specularity values in certain areas is due to finite acquistion

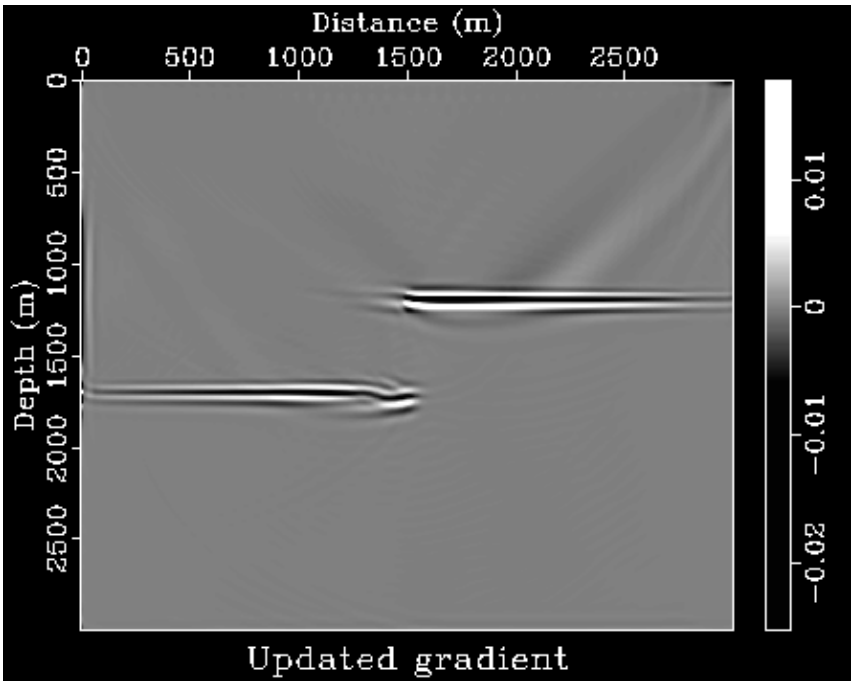


Figure 5.9: Gradient

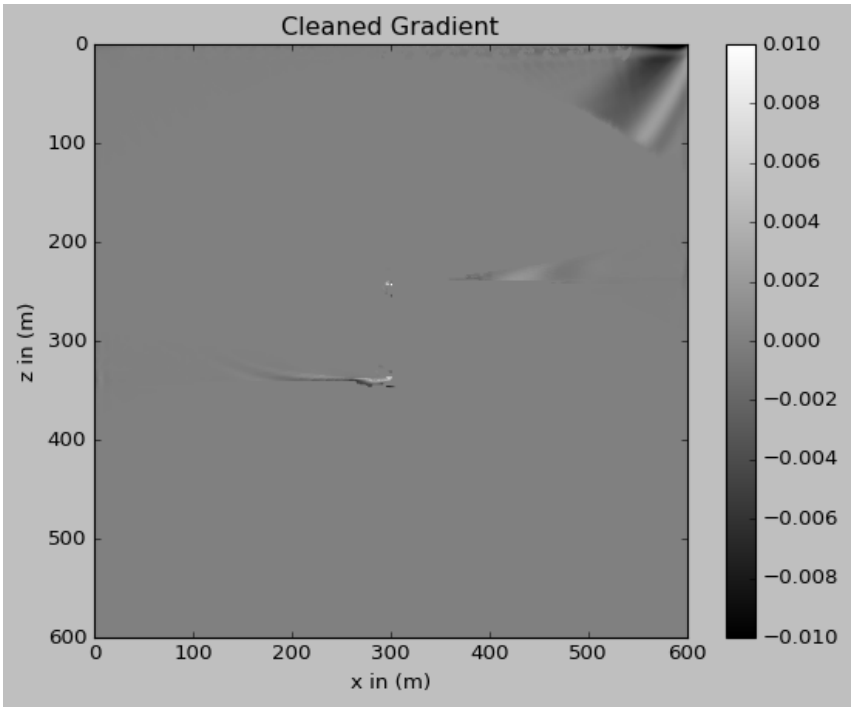


Figure 5.10: Cleaned Gradient after 1^{st} iteration

Example-2 :Diffraction Ramp Model

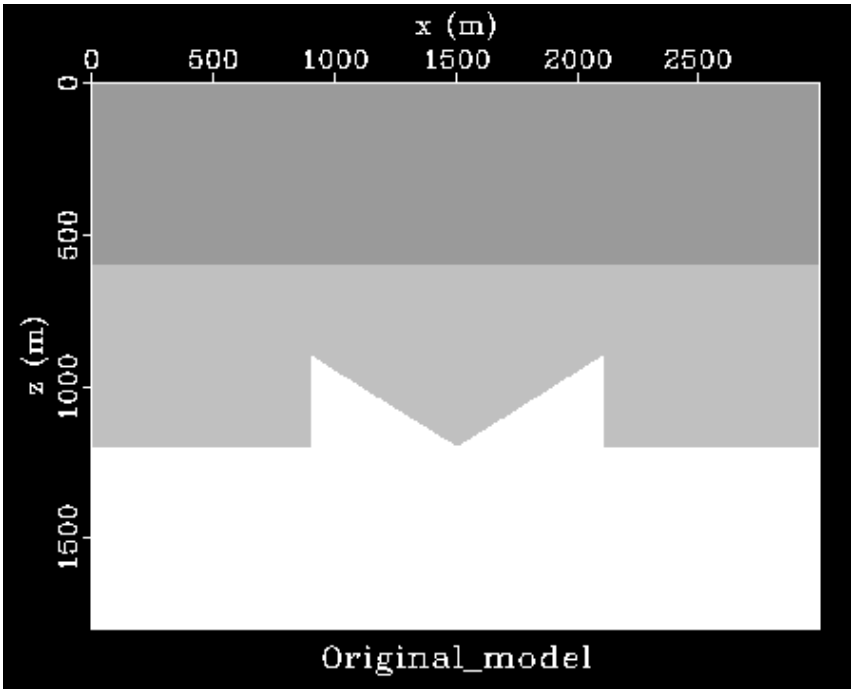


Figure 5.11: Velocity model

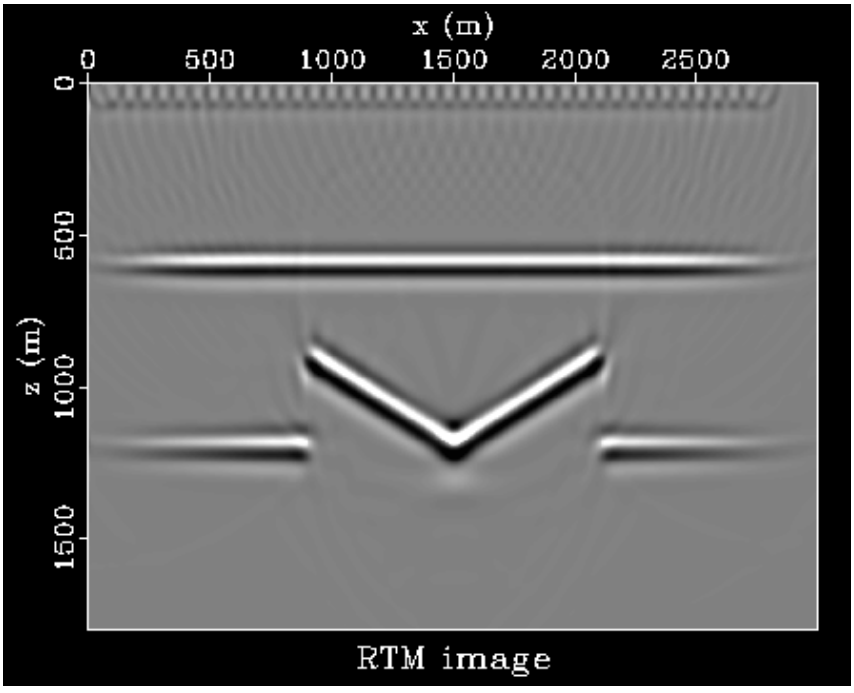


Figure 5.12: RTM image

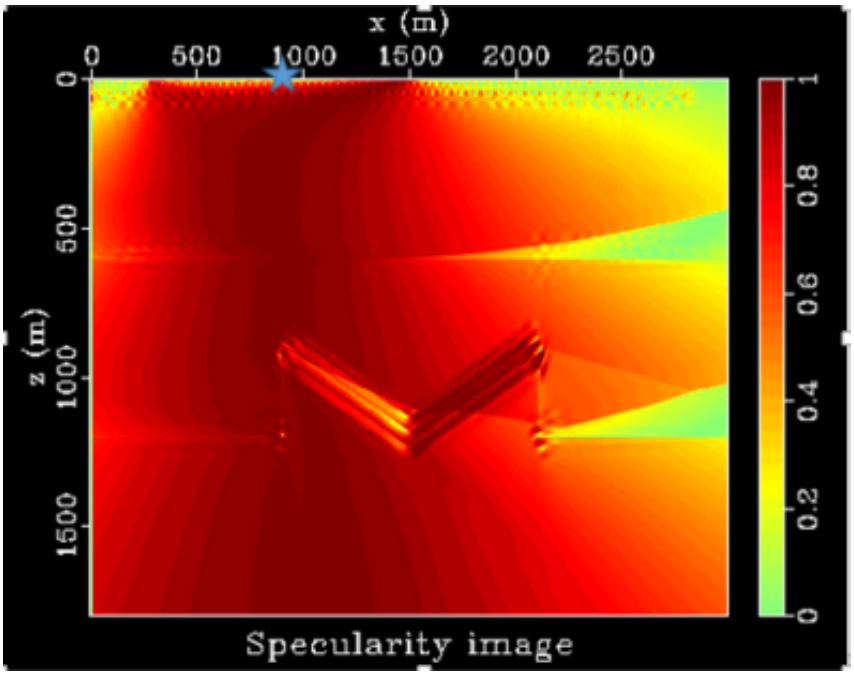


Figure 5.13: Specularity image

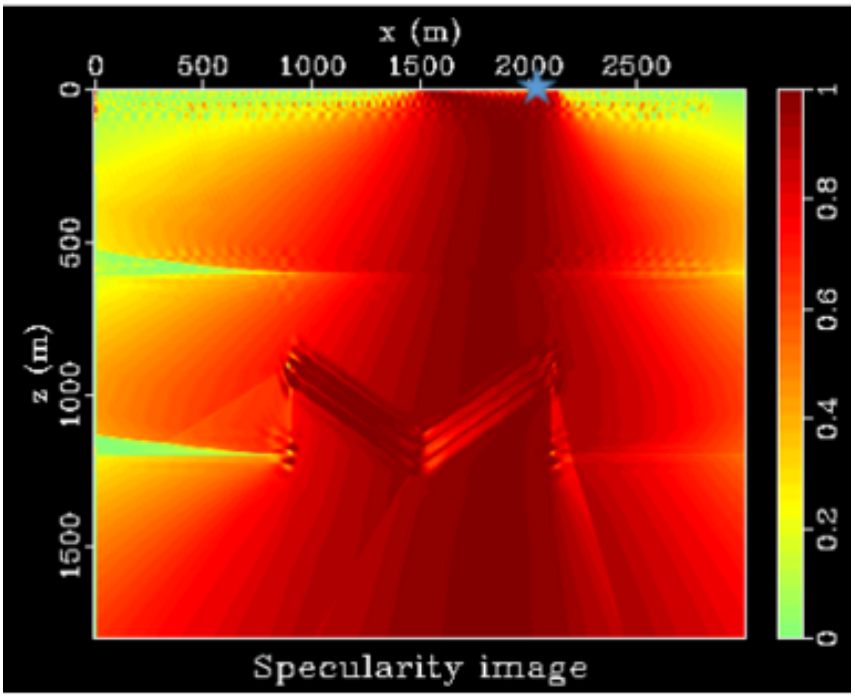


Figure 5.14: Specularity image

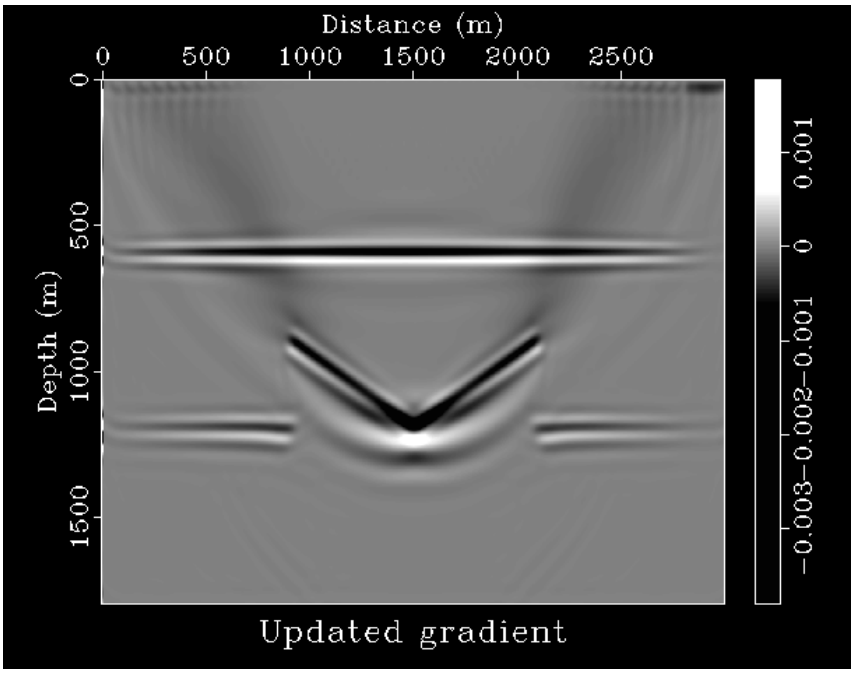


Figure 5.15: Gradient

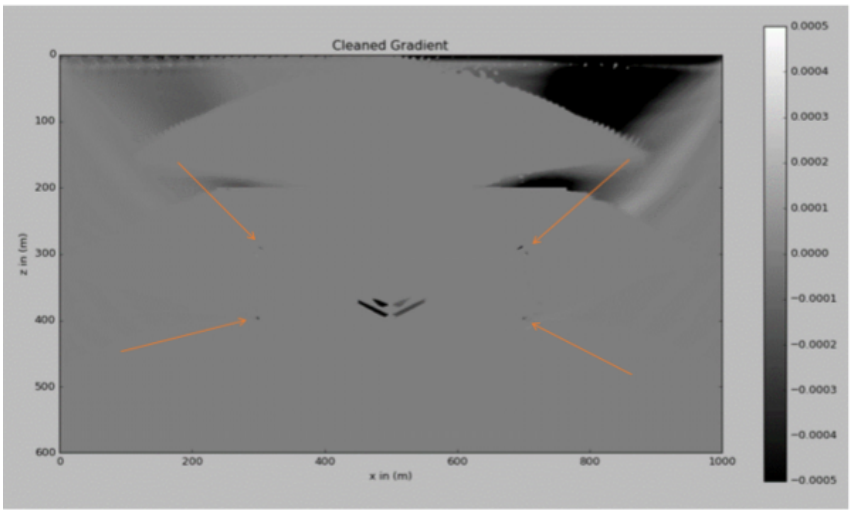


Figure 5.16: Cleaned Gradient after 1^{st} iteration

5.0.9 Wave based Specularity

In this section we will try to extend the definition of specularity measure to wavefields . We define specularity as

$$S(s,x,r) = \frac{\left|n^P.(P_S + P_R)\right|}{\|P_S + P_R\|} \tag{5.4}$$

where P is direction of poynting vector. Poynting vector is a measure of energy flow and is defined as

$$P = -\nabla u \frac{\partial u}{\partial t} \tag{5.5}$$

where P_S is the direction of source wavefield gradient and P_R is the direction of receiver wavefield gradient. The sum of these vectors should be nearly aligned with normal vector for specular reflection. We attempt to utilise this definition in directly cleaning the gradient of Full Waveform Inversion. I used time averaged poynting vectors to stabilise the direction and thus for cleaning Poynting vectors could be easily calculated in the back-propagation step so the cost involved in calculating the poynting vectors is very low.

Example-1 : Step Layer Model

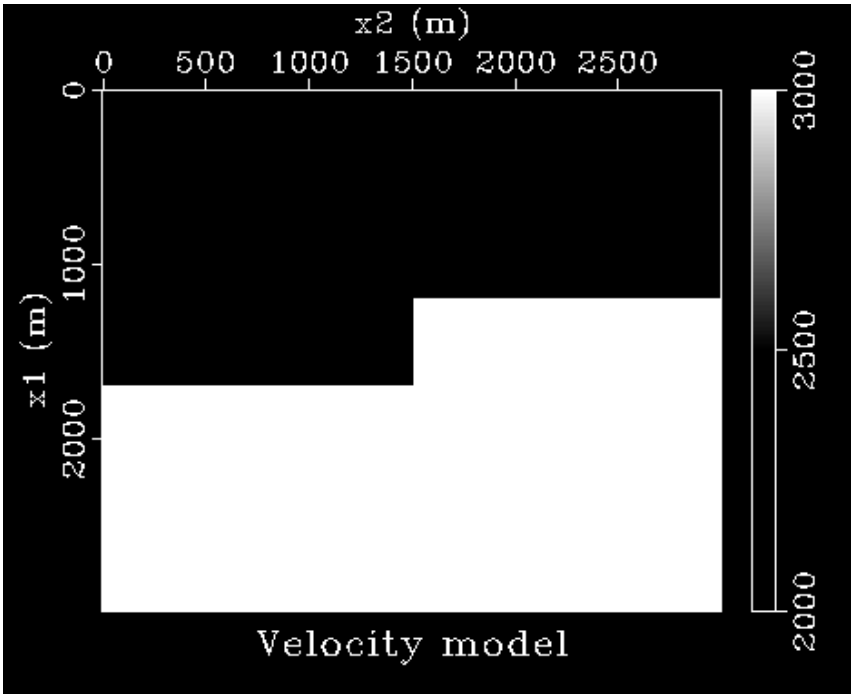


Figure 5.17: Velocity model

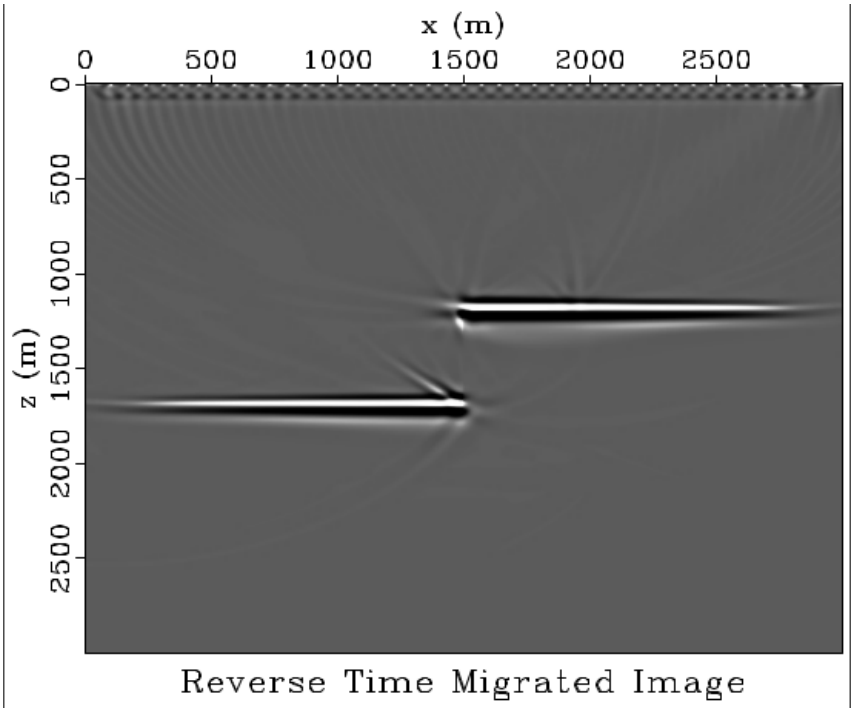


Figure 5.18: RTM image

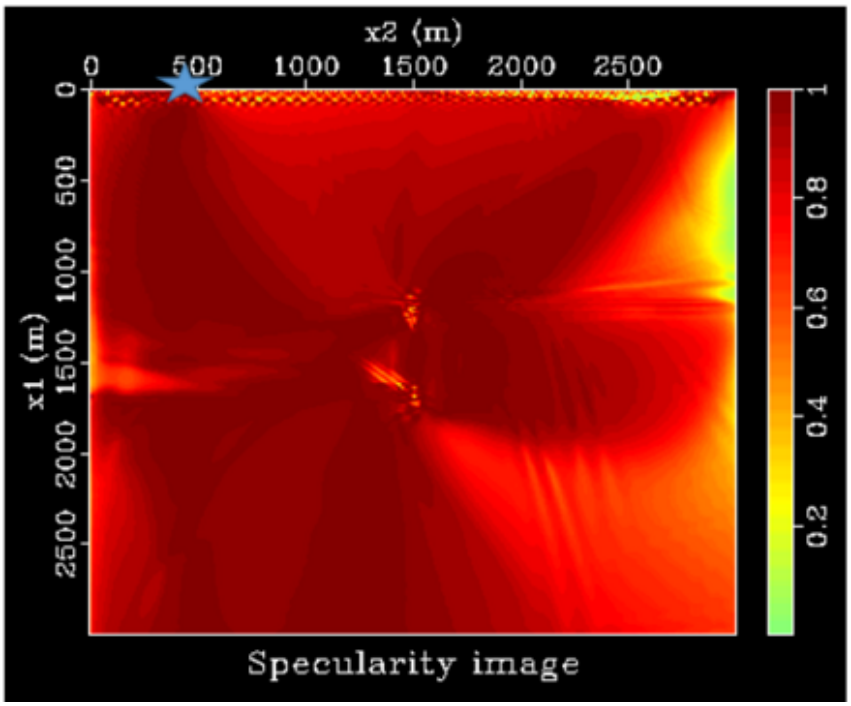


Figure 5.19: Specularity image

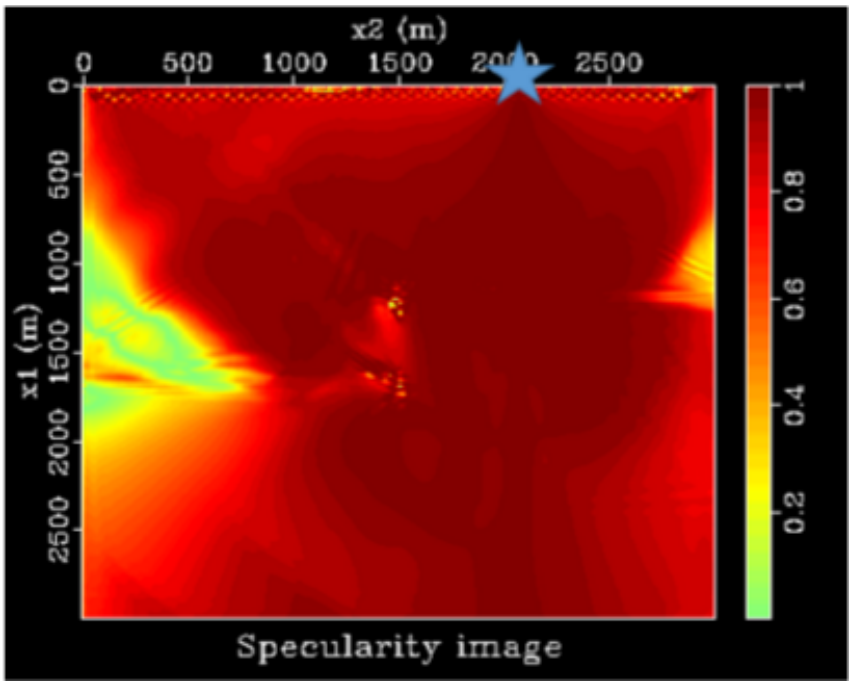


Figure 5.20: Specularity image

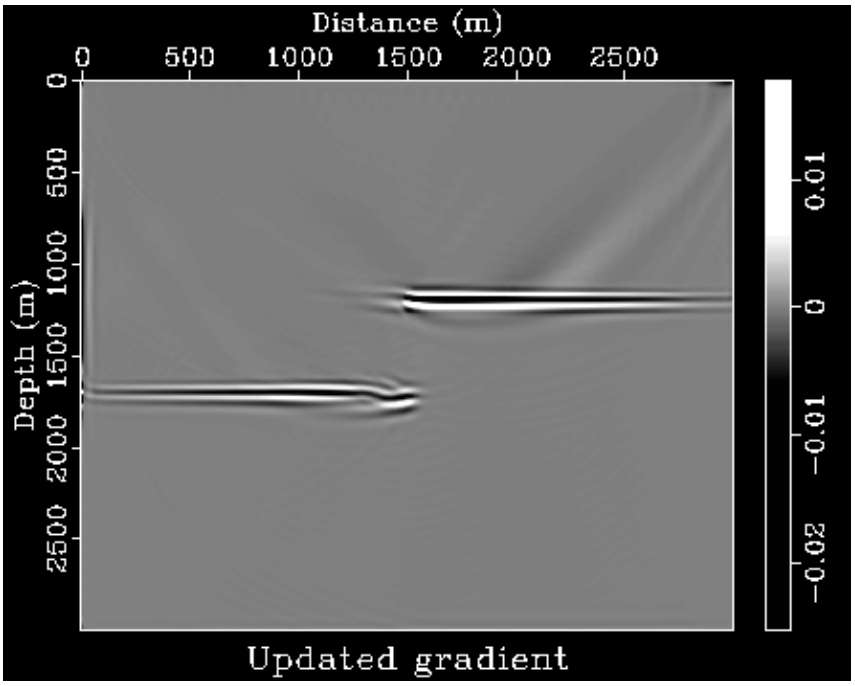


Figure 5.21: Gradient

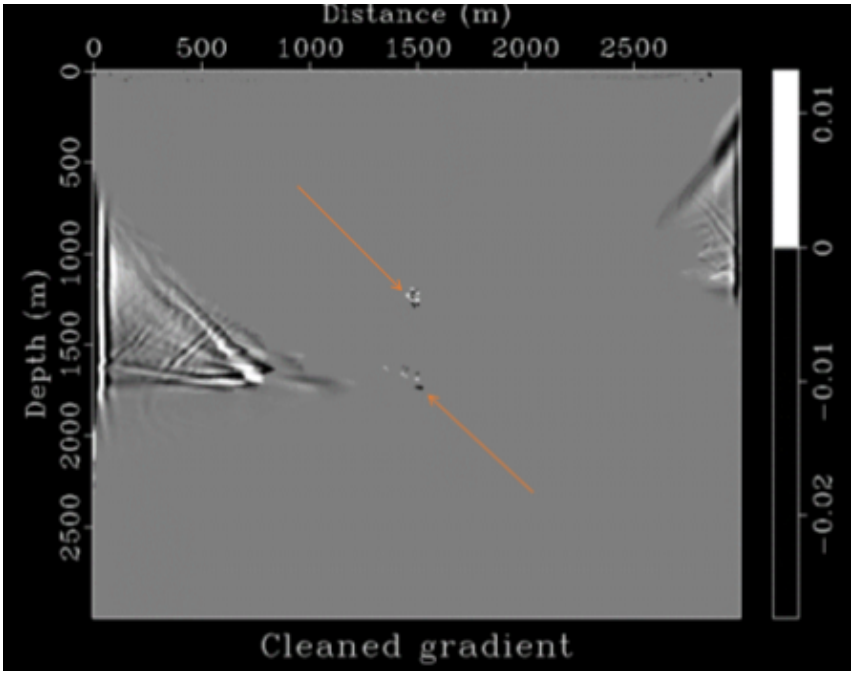


Figure 5.22: Cleaned Gradient after 1^{st} iteration

Example-2 :Diffraction Ramp Model

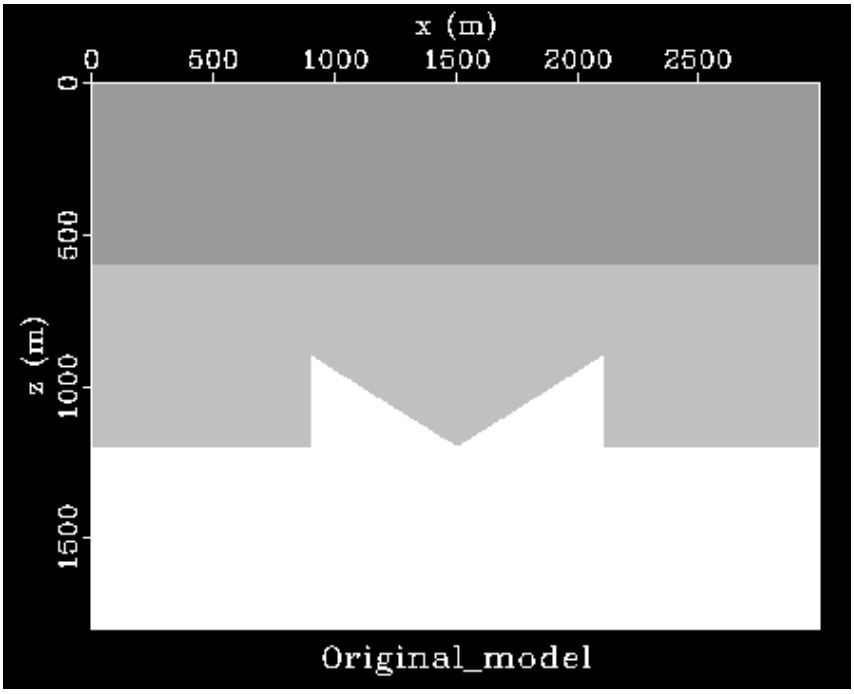


Figure 5.23: Velocity model

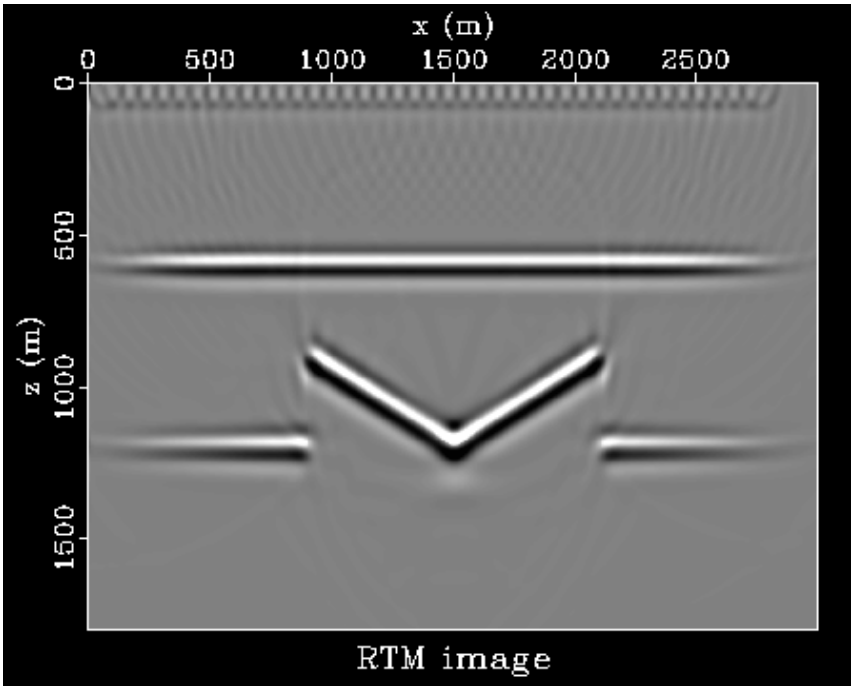


Figure 5.24: RTM image

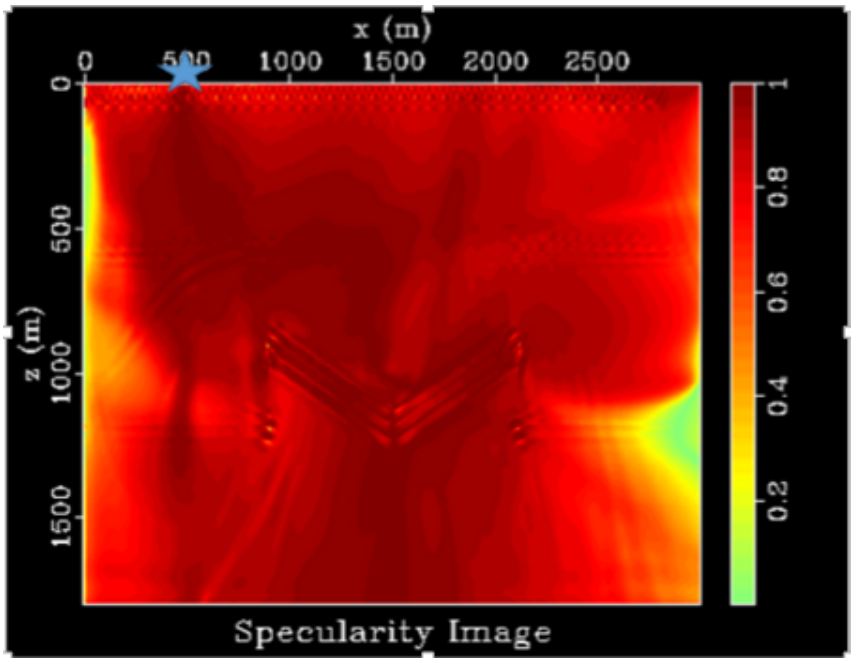


Figure 5.25: Specularity image

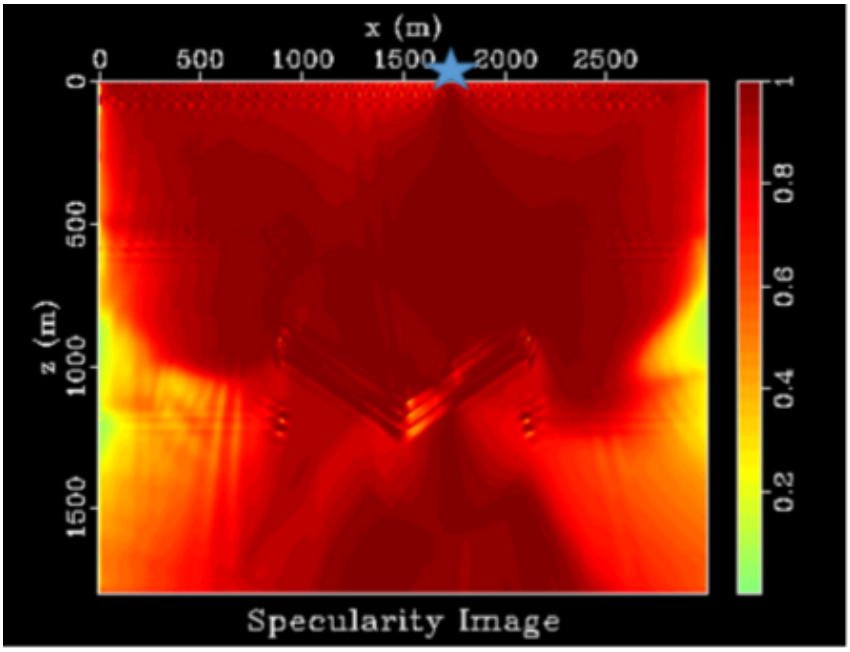


Figure 5.26: Specularity image

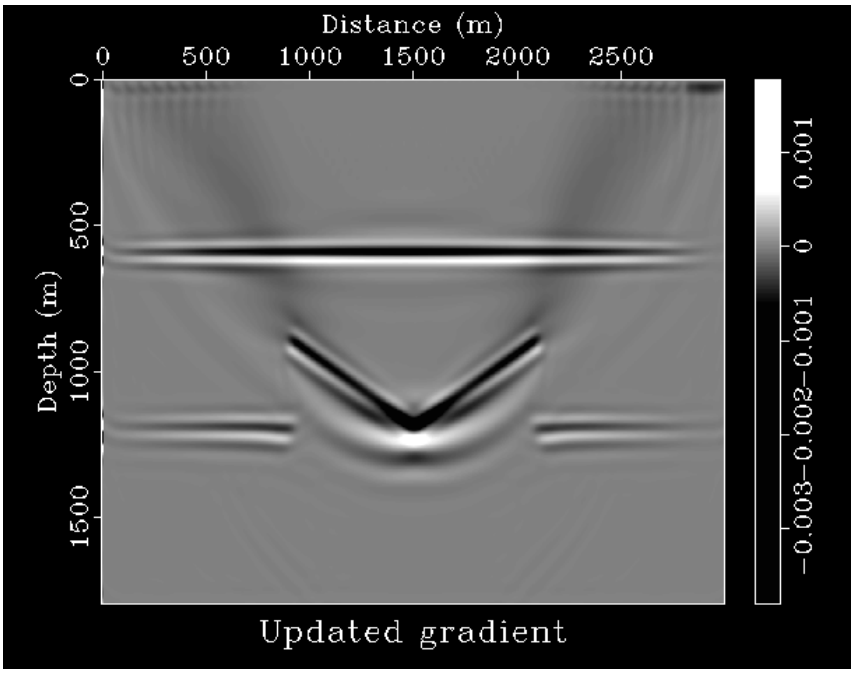


Figure 5.27: Gradient

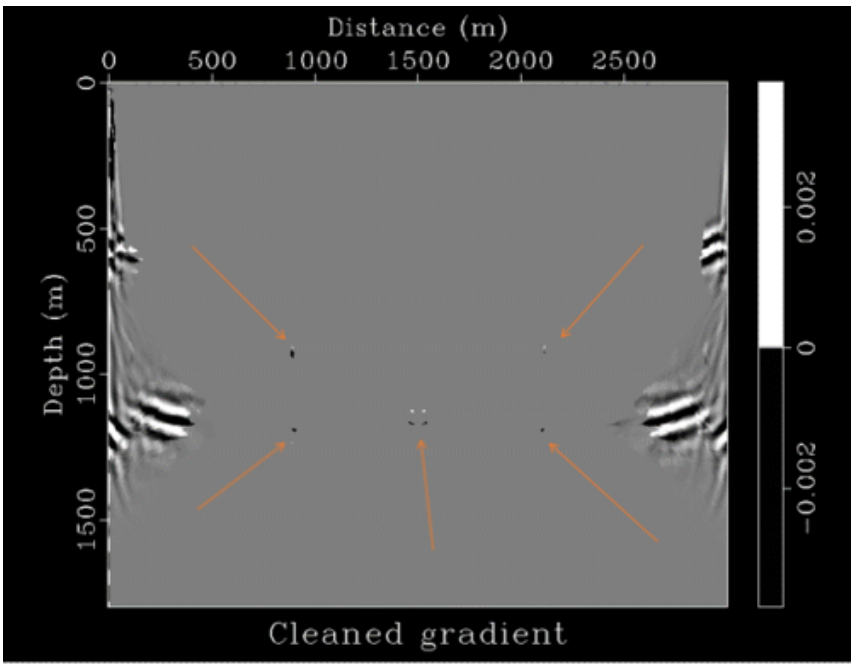


Figure 5.28: Cleaned Gradient after 1^{st} iteration

Example-3 :Marmousi Model with Diffractors

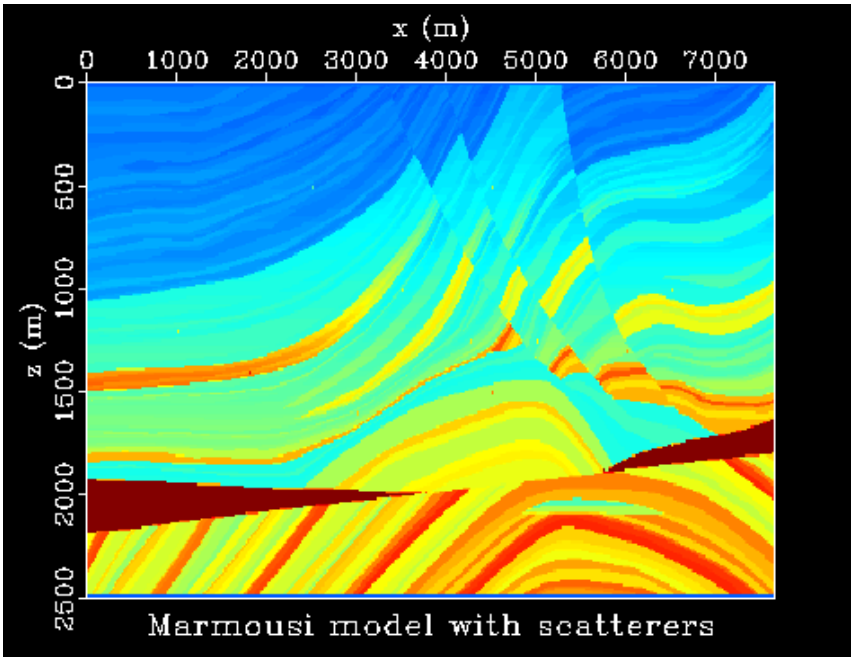


Figure 5.29: Velocity model

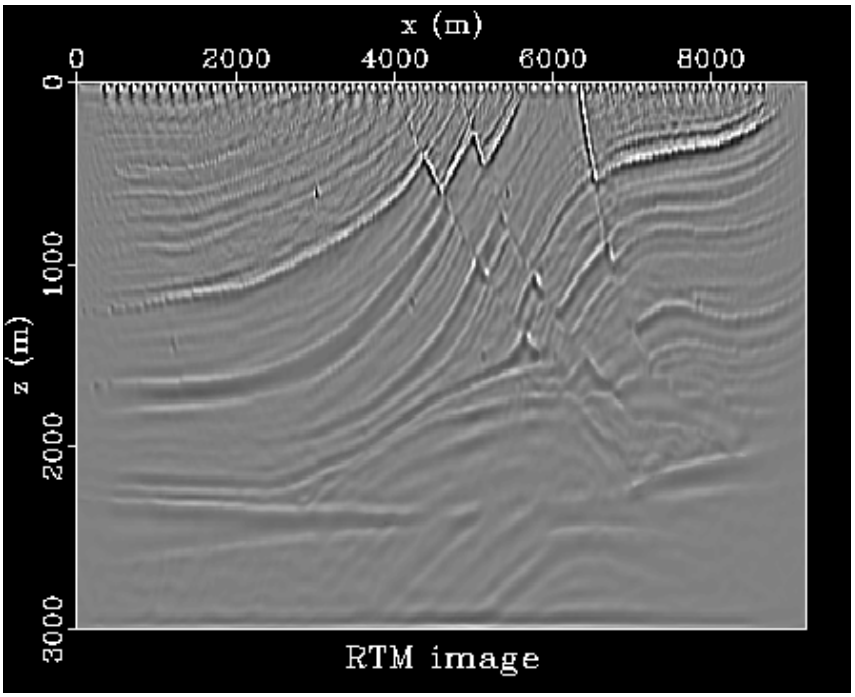


Figure 5.30: RTM image

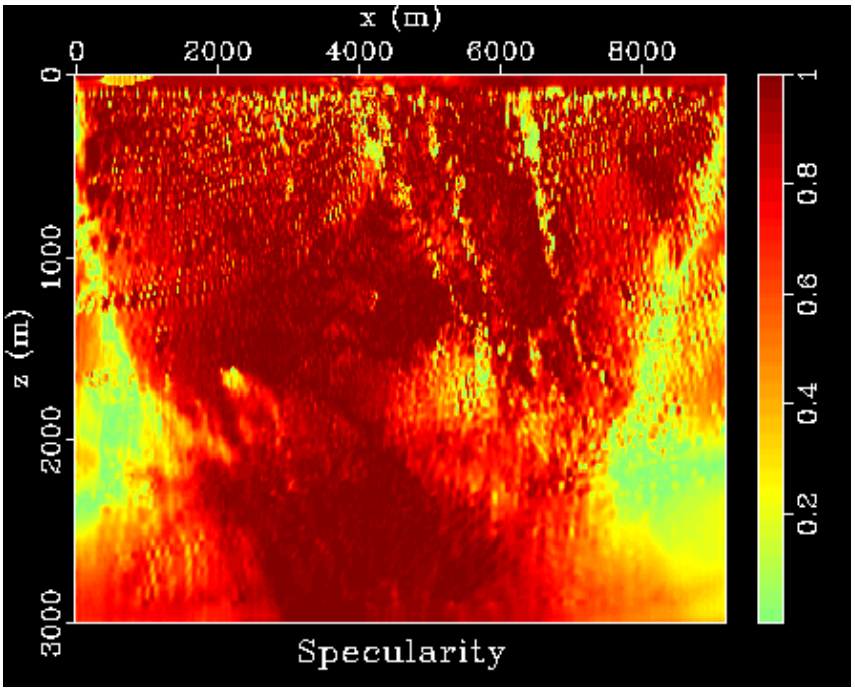


Figure 5.31: Specularity image

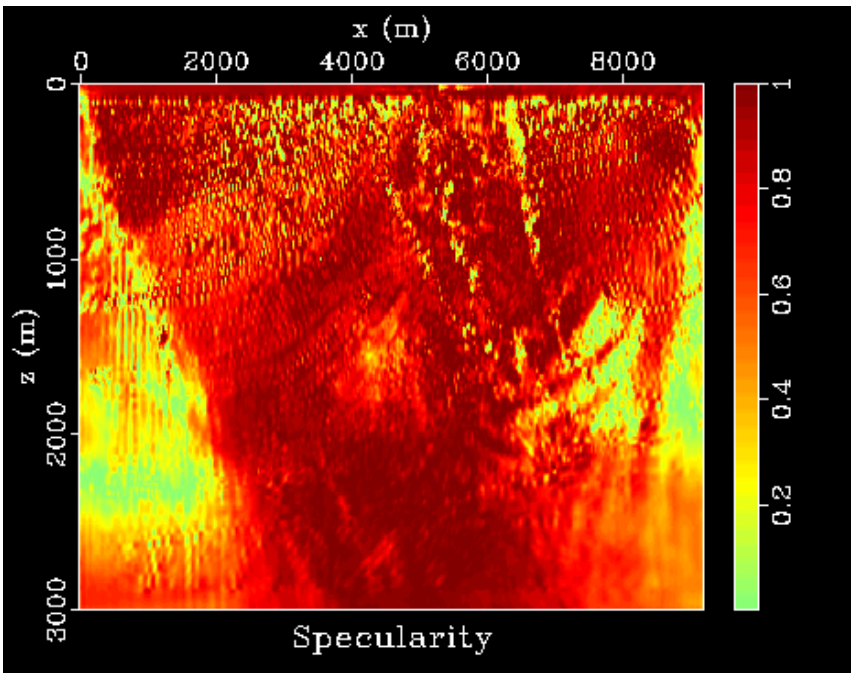


Figure 5.32: Specularity image

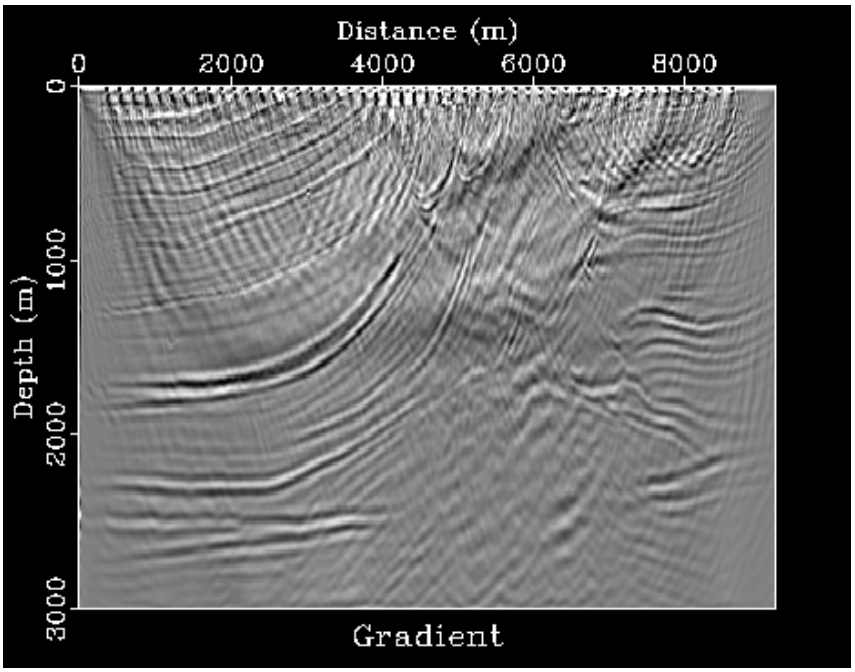


Figure 5.33: Gradient

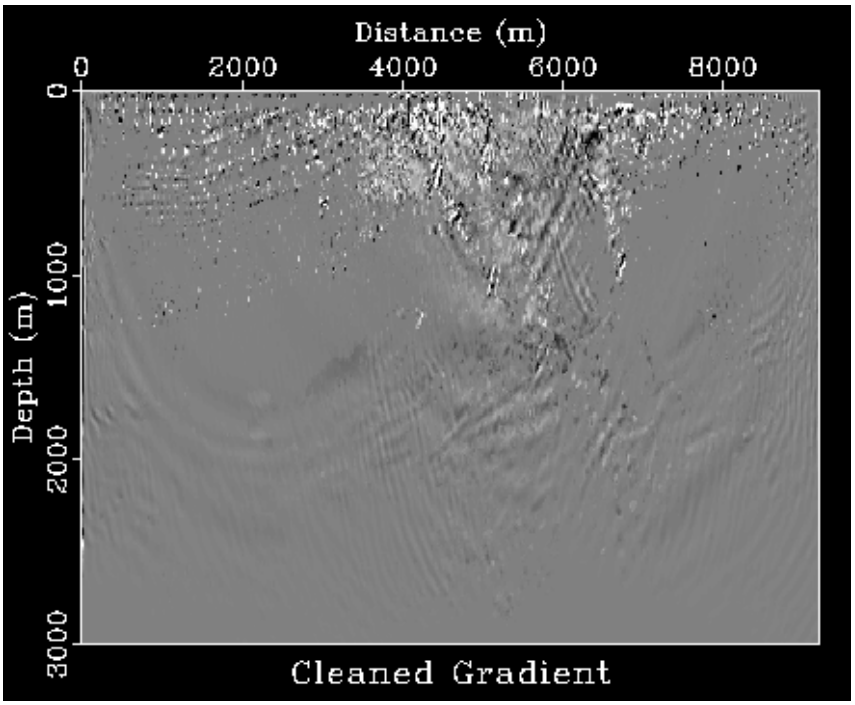


Figure 5.34: Cleaned Gradient after 1^{st} iteration

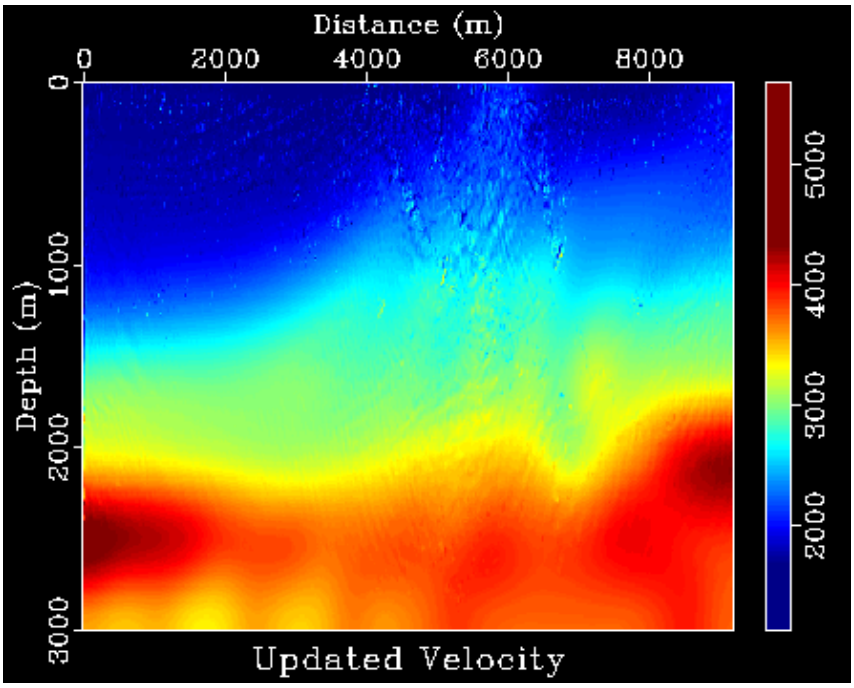


Figure 5.35: Velocity Model after 60_{th} iteration

5.1 Discussion

For the specularity computation a good velocity model is required since our aim is to obtain the high resolution information corresponding to diffractors . We only go for diffraction imaging if our reflection events are imaged properly . Thus we showed that it is possible to pin-point the diffractors which could be geologically related to small scale features like pinchouts , faults etc. However specularity cleaning also worked if our velocity model is far from slighlty smooth model but the criteria is to maintain the kinematics of the model otherwise the diffraction updates will be defocussed over a larger region . The updates still get impressions of diffraction but they doesn't correspond to high resolution information what we are aiming for . For a step layer case instead of slightly smoothing we started with a constant velocity background with velocity equal to the top layer of the velocity to honor the kinematics . The velocities only jumped by $200m/s$ after 20 iterations as it can't jump directly to the true velocity however the updates were confined

to our original diffractor location.

Chapter 6

Diffraction Inversion

In previous section we try to clean inherently the specular part to obtain the contibution from the diffractors . In this section we will attempt to separate or predict diffractions using the Born framework and then we will try to invert solely the diffractions using Full Waveform Inversion

6.0.1 Scattering series and Born approximation

Scattering series and linearization are the basis of most inversion methods , both direct and iterative. In seismic it is feasible to separate our medium into smooth background part and rough perturbed part.We explain the scattered waves as reflections and diffractions due to rough perturbed part to the background. We write

$$m(x) = \frac{1}{v(x)^2}, \quad and \quad m_0(x) = \frac{1}{v_0(x)^2} \tag{6.1}$$

for a small number α

$$m(x) = m_0(x) + \alpha m_1(x) \tag{6.2}$$

Note that when perturbing velocity additional Taylor series expansion is necessary.

$$v(x) = v_0(x) + \alpha v_1(x) \quad \rightarrow \quad \frac{1}{v(x)^2} = \frac{1}{v_0(x)^2} - \frac{2\alpha v_1(x)}{v_0(x)^3} \tag{6.3}$$

The wave equation for the acoustic media can be written as

$$m(x)\frac{\partial^2 u(x,t)}{\partial t^2} - \nabla^2 u(x,t) = f(x,t) \tag{6.4}$$

The corresponding wavefield will split into

$$u(x,t) = u_0(x,t) + u_{sc}(x,t) \tag{6.5}$$

where $u_0(x,t)$ will solve the wave equation for undisturbed medium $m_0(x)$

$$m_0(x)\frac{\partial^2 u_0(x,t)}{\partial t^2} - \nabla^2 u_0(x,t) = f(x,t) \tag{6.6}$$

$u(x,t)$ is the total wavefield , $u_0(x,t)$ is the incident wavefield and $u_{sc}(x,t)$ is the scattered wavefield (anything but the incident wavefield) we get the equation for $u_{sc}(x,t)$ by subtracting 128 by 127 .

$$m_0(x)\frac{\partial^2 u_{sc}(x,t)}{\partial t^2} - \nabla^2 u_{sc}(x,t) = -\alpha m_1(x)\frac{\partial^2 u(x,t)}{\partial t^2} \tag{6.7}$$

The equation is implicit in sense that the right hand still depends on the $u_{sc}(x,t)$ through $u(x,t)$. We can reformulate this implicit integral by the means of Green's function

$$u_{sc}(x,t) = -\alpha \int_0^t \int_R G(x,y,t-s)m_1(y)\frac{\partial^2 u(y,s)}{\partial t^2} dyds \tag{6.8}$$

we write

$$u_{sc} = -\alpha G m_1 \frac{\partial^2 u}{\partial t^2} \tag{6.9}$$

where G is the operator for space-time integration . In terms of u we have an implicit relation

$$u = u_0 - \alpha G m_1 \frac{\partial^2 u}{\partial t^2} \tag{6.10}$$

called a $Lipman - Schwinger$ equation. u can be witten as

$$u = \Big[I + \alpha G m_1 \frac{\partial^2}{\partial t^2}\Big]^{-1} u_0 \tag{6.11}$$

Writing the above $[I+A]^{-1}$ for some operator invites the solution in the form of Neumann series $I - A + A^2 - A^3 + ...$ provided $||A|| < 1$ in some norm . We rewrite the equation as

$$u = u_0 - \alpha\Big(G m_1 \frac{\partial^2}{\partial t^2}\Big) + \alpha^2\Big(G m_1 \frac{\partial^2}{\partial t^2}\Big)\Big(G m_1 \frac{\partial^2}{\partial t^2}\Big) + .. \tag{6.12}$$

The above series is called Born series based on the weak scattering condition $\alpha||Gm_1\frac{\partial^2}{\partial t^2}|| < 1$. It carries the physics of multiple scattering.

$$u_0 \quad - \alpha \int_0^t \int_R G(x,y,t-s)m_1(y)\frac{\partial^2 u_0(y,s)}{\partial t^2}dyds + ... \tag{6.13}$$

$$(incidentwavefield) \qquad\qquad (singlescattering) \tag{6.14}$$

$$\tag{6.15}$$

summarizing the expansion as

$$u = u_0 + \alpha u_1 + \alpha^2 u_2 + .. \tag{6.16}$$

where αu_1 represent single scattering,$\alpha^2 u_2$ double scattering, etc. For instance, the expression of u_1 can be physically read as "the incident wave initiates from the source at time $t = 0$, propagates to y where it scatters due to $m(y)$ at time $t = s$, then further propagates to reach x at time t.

The expression of u_2 can be read as "the incident wave initiates from the source at $t = 0$, propagates to y_1 where it first scatters at time $t = s_1$, them propagates to y_2 where it scatters a second time at time $t = s_2$, then propagates to x at time t, where it is observed. The approximation $u_sc = \alpha u_1$ is called Born approximtion. From this we can return to our PDE.

$$m_0(x)\frac{\partial^2 u_{sc}(x,t)}{\partial t^2} - \nabla^2 u_{sc}(x,t) = -\alpha m_1(x)\frac{\partial^2 u_0(x,t)}{\partial t^2} \tag{6.17}$$

The only difference is the presence of u_0 in place of u in the right hand side (and α is gone, by choice of normalization of u_1) .The weak scattering condition $\alpha||u_1||' < ||u_0||'$ encodes the idea that the primary reflected field $\alpha u_1$is small compared to the incident field u_0 . It is satisfied when α is small, and when u_1 is not so large that it would undo the smallness of α.

The Born approximation is often satisfied in the low-frequency regime, by virtue of the fact that cycle skipping is not as much of an issue. In the high-frequency regime, the heuristics for validity of the Born approximation are that

1) v_0 or m_0 should be smooth.

2) v_1 or m_1 should be localized, or better yet, localized and oscillatory (zero mean).

6.0.2 Diffraction Free model

The idea of diffraction free model is to somehow isolate the diffractions in the model domain. There are numerous ideas through which this feat can be performed , some of them include utilising Structurally oriented tensors , wavelet transforms , Dip steering filters etc . We will utilise a dip steering filter to separate the diffractions . The dip steering filter is a data-driven tool (in model domain our data is model parameters values) that enhances the coherent part which will corresponds to reflection and suppress the abrupt values which generates the diffractions.Th filter collects all amplitudes inside a disc with user-specified search radius and replaces the value at the center by an avergae value of the amplitudes. The search disk then follows the local dips and azimuth along with steering cube.

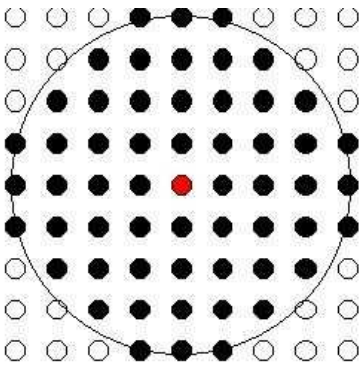


Figure 6.1: Dip steering radius or width (source:www.opendtect.com)

6.0.3 Numerical Results: Diffraction Inversion

The workflow is to separate diffractivity model from the original model using dip steering filter and model diffractions using Born modeling . Then we will invert using Full Waveform Inversion using Diffractions as our observed data.

6.0.3.0.4 Point Scatterer test :Born modeling

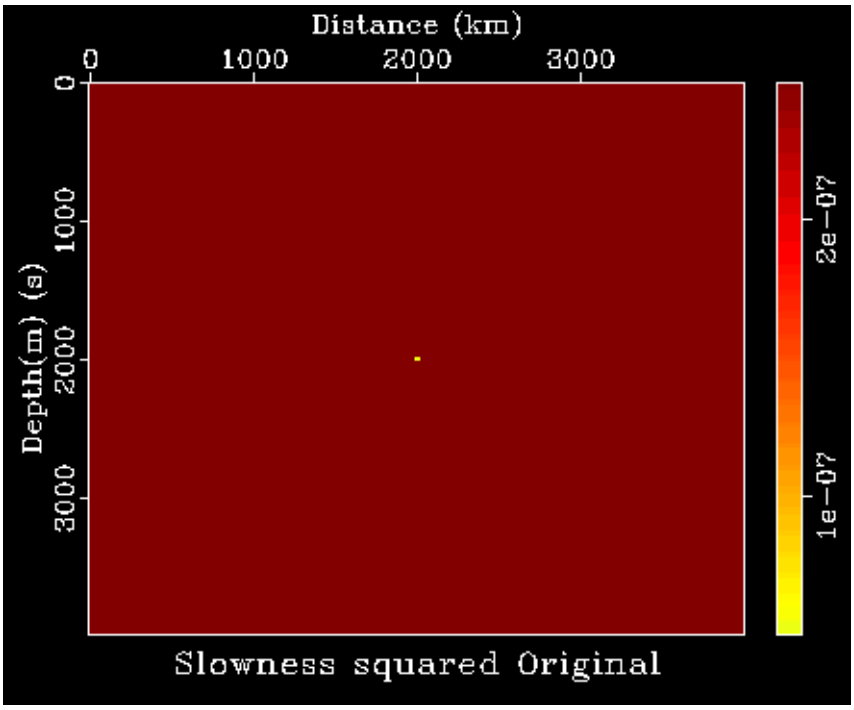


Figure 6.2: Point Scatterer model

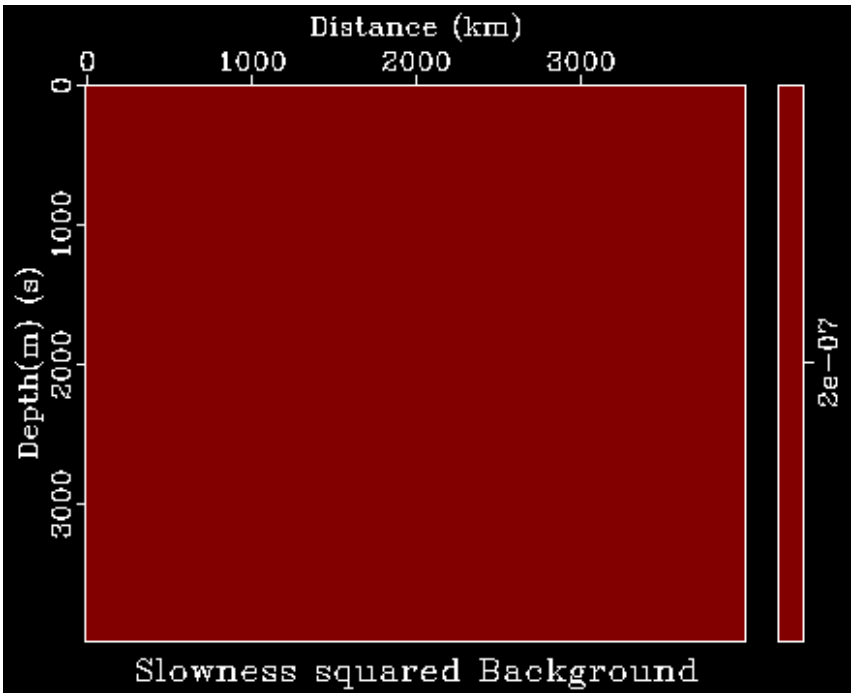


Figure 6.3: Background Model

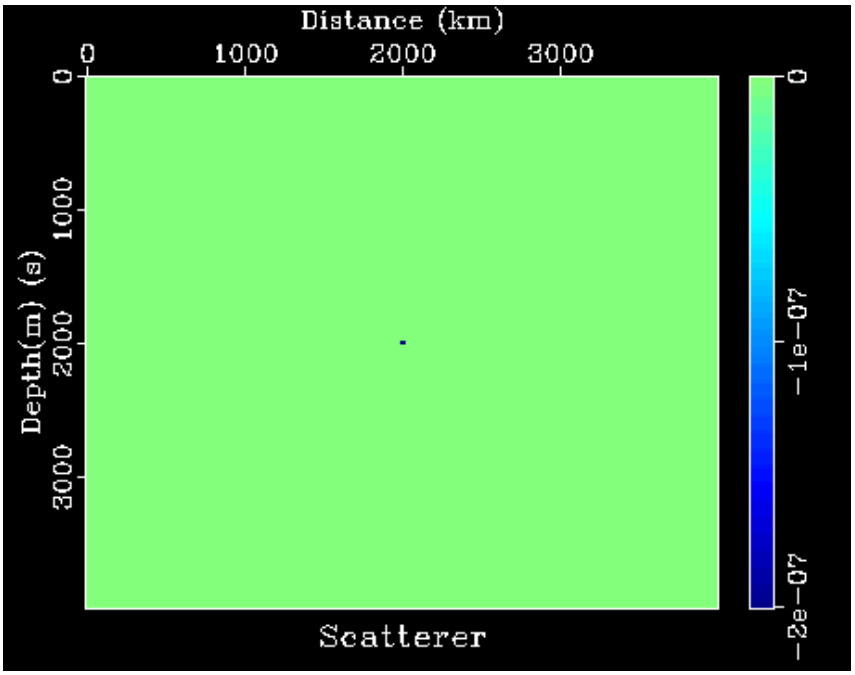


Figure 6.4: Perturbed Model

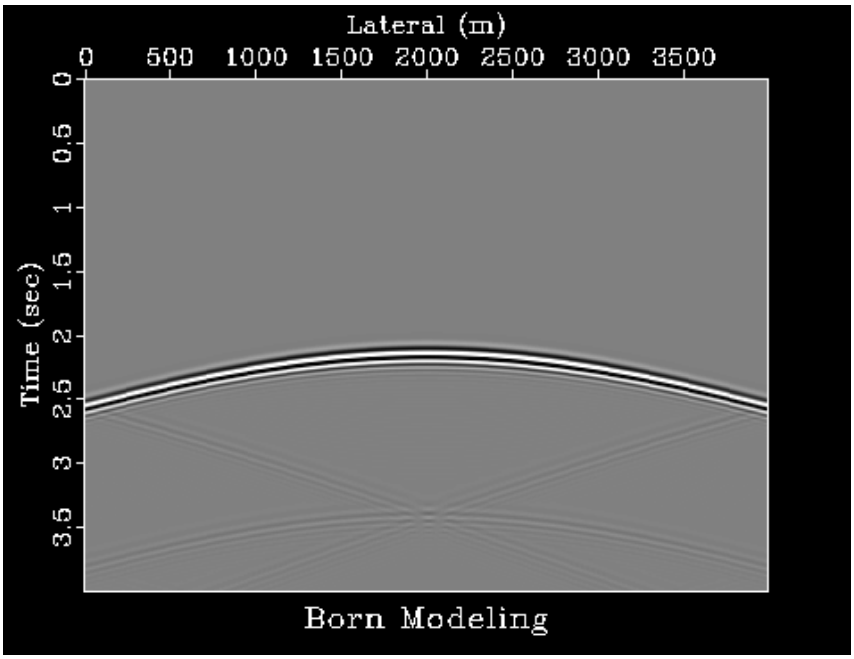


Figure 6.5: Diffractions

6.0.3.0.5 Marmousi DFM test Marmousi Model with Induced Scatterers :

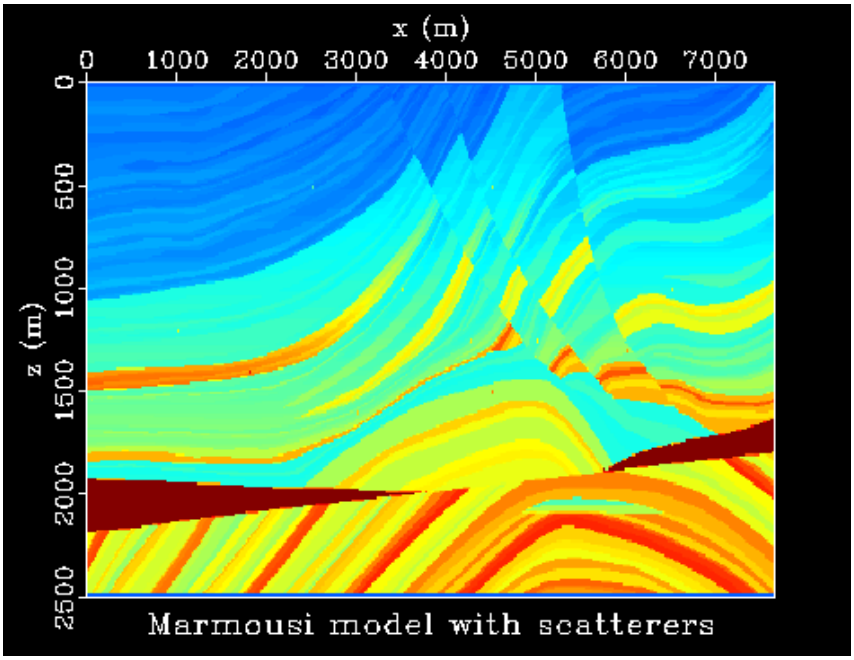


Figure 6.6: Mamrousi model with scatterers

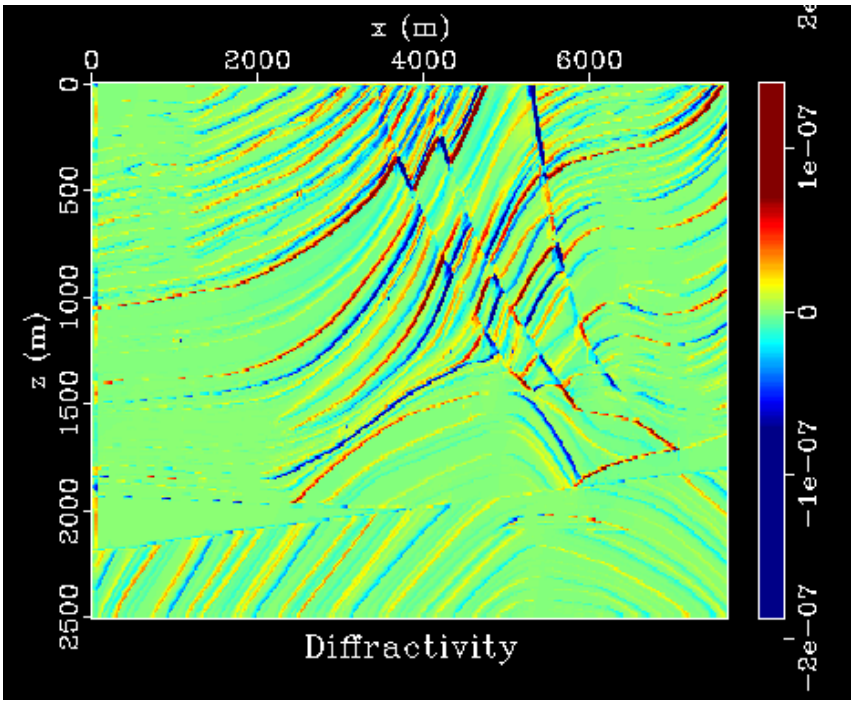


Figure 6.7: Diffractivity Model

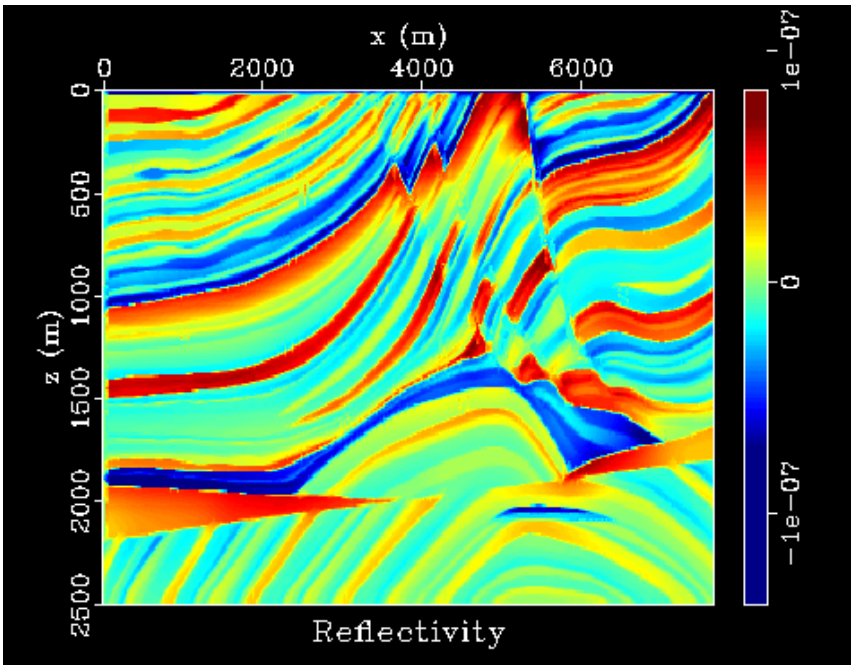


Figure 6.8: Reflectivity Model

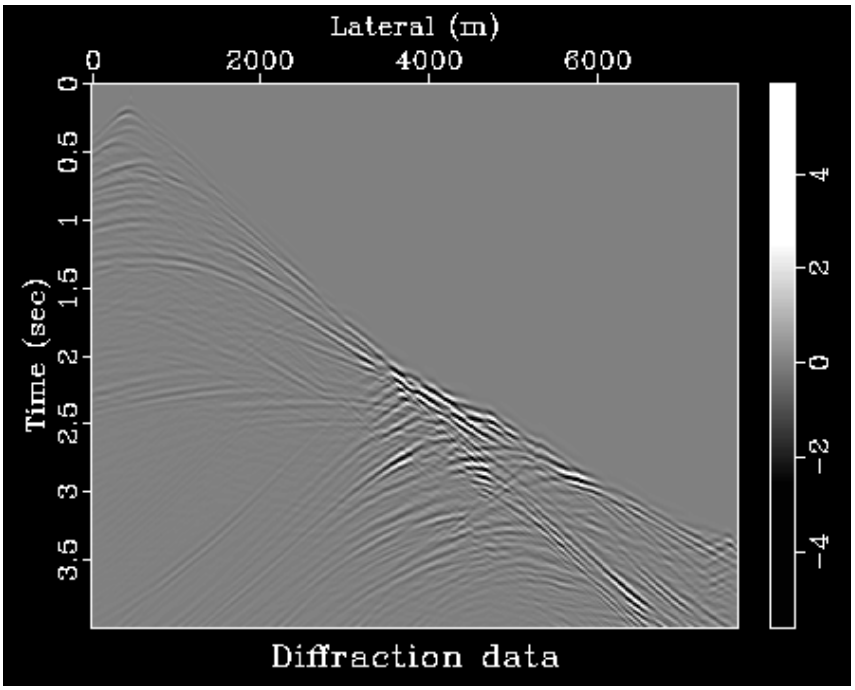


Figure 6.9: Diffractions

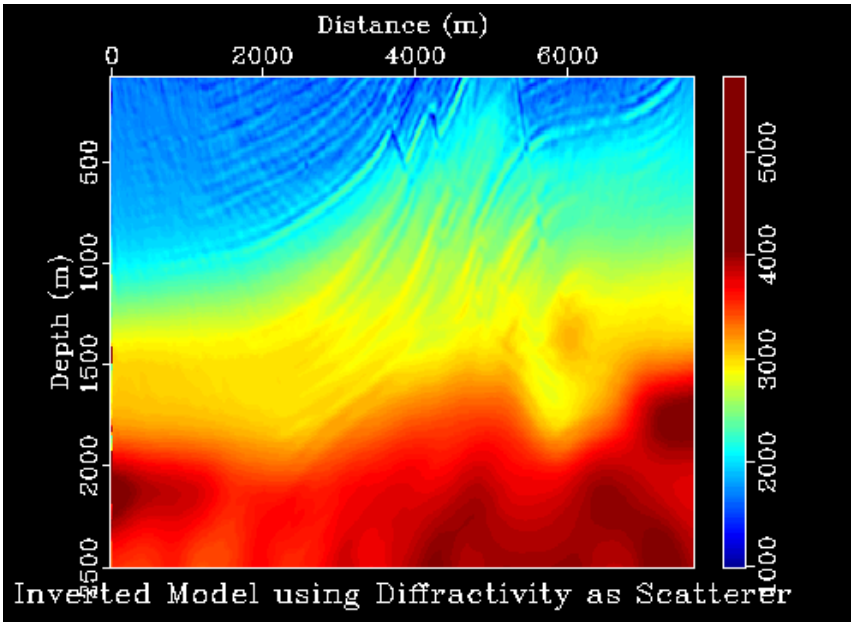


Figure 6.10: Inverted Model using Diffractions

6.1 Conclusion and Future Work

We here showed that Diffractions could be inverted separately using Full Waveform Inversion . However the idea is to further develop Migration Velocity Analysis approaches for Diffraction where we believe that Full Waveform Inversion could updating the long-wavelength velocity variations and focusing the diffractions which requires further more work . The plan is to further look into conjecture of Prof Tariq looking into Reflection Full Waveform Inversion.

Appendices

Appendix A

The Adjoint State Method

I employed adjoint state method for the computation of gradient term in Full Waveform Inversion . Here I will introduce the general theory of adjoint state method which is not limited to wave propagation problems and I also the address the situations where it is in general applicable.

Let's say if you have a functional

$$J(m) = h(u(m), m) \tag{A.1}$$

and you want to get $\frac{\partial J}{\partial m}$. If we take derivative w.r.t m we get

$$\frac{\partial J}{\partial m} = \frac{\partial h}{\partial u}\frac{\partial u}{\partial m} + \frac{\partial h}{\partial m} \tag{A.2}$$

If the term $\frac{\partial u}{\partial m}$ is quite expensive to compute and you want to somehow avoid computing it there the adjoint state methods comes in. Let's relate this to our situation

Here functional $J(m)$ is equivalent to our L_2 norm objective function in Full Waveform Inversion and $\frac{\partial u}{\partial m}$ is equivalent to explicitly computing the Frechet derivative or Jacobian which is quite expensive. The adjoint state method states (Plessix [**?**]) building a new augmented (Lagrangian) functional and using this to define adjoint state equations which lead to the computation of $\frac{\partial J}{\partial m}$. We will use our original functional $J(m)$ to build a Lagrangian functional and usually the $u(m)$ is constrained by some relation $F(u(m), m) =$

0 (equivalent to forward modeling operator $Am = u$)

We define our augmented functional as

$$L(\tilde{u}, \tilde{\lambda}, m) = h(\tilde{u}(m), m) - < \tilde{\lambda}, F(\tilde{u}(m), m) > \tag{A.3}$$

where $\tilde{u}$ is the generalized state variable and $\tilde{\lambda}$ is the adjoint state variable that both independent of m .Now we replace the generalized state variable $\tilde{u}$ with our physical realized variable u and as we know that u follows the relation $F(u, m) = 0$ so the equation reduces to

$$L(\tilde{u}, \tilde{\lambda}, m) = h(\tilde{u}, m) \tag{A.4}$$

From equation 140 , we get

$$J(m) = L(u, \tilde{\lambda}, m) \tag{A.5}$$

Now we take derivative w.r.t m

$$\frac{\partial J}{\partial m} = \frac{\partial L(u, \tilde{\lambda}, m)}{\partial m} + \frac{\partial L(u, \tilde{\lambda}, m)}{\partial u}\frac{\partial u}{\partial m} \tag{A.6}$$

We note we still have the term $\frac{\partial u}{\partial m}$ but this time the difference is our functional depends upon λ (independent of m) we can choose it such that the term $\frac{\partial L(u,\tilde{\lambda},m)}{\partial u}$ is zero .

$$\frac{\partial L(u, \tilde{\lambda}, m)}{\partial u} = \frac{\partial h(u, m)}{\partial u} - \Big(\frac{\partial F(u, m)}{\partial u}\Big)\tilde{\lambda} = 0 \tag{A.7}$$

we solve for the above equation for $\tilde{\lambda}$ to get the adjoint state equations and then put back into equation 145.

$$\frac{\partial J}{\partial m} = \frac{\partial L(u, \tilde{\lambda}, m)}{\partial m} \tag{A.8}$$

Now using our original definition of $L(u, \tilde{\lambda}, m)$, we get

$$\frac{\partial J}{\partial m} = \frac{\partial h(u, m)}{\partial m} - \Big(\frac{\partial F(u, m)}{\partial m}\Big)^{*}\tilde{\lambda} \tag{A.9}$$

A.0.0.0.6 Time Domain Gradient computation : In time domain our wave operator $L = \left(\frac{1}{v(x)^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)$ satisfies following equation

$$Lu = f \tag{A.10}$$

with initial boundary conditions $u(0) = 0, \frac{\partial u}{\partial t}|_{t=0} = 0$. The objective function in time domain is

$$\frac{1}{2}\int_{t=0}^{t=T}\sum_{s=1}^{s=N_s}\sum_{r=1}^{s=N_r}||P_r u_{obs}(t) - d_{pre}(t)||_2^2 \tag{A.11}$$

where P is a projection operator which model space to receiver space. The augmented functional is defined by

$$L(u,\tilde{\lambda},m) = \frac{1}{2}\int_{t=0}^{t=T}\sum_{s,r}||P_r u_{obs}(t) - d_{pre}(t)||_2^2 - \sum_s\int_0^T\Big\langle\tilde{\lambda}, m\frac{\partial^2 u}{\partial t^2} - \nabla^2 u - f_s\Big\rangle dt \tag{A.12}$$

with $<\tilde{\lambda}, u>=\int_x \tilde{\lambda}udx$ scalar product in real coordinate space.Now using integration by parts twice

$$\int_0^T\Big\langle\tilde{\lambda}, m\frac{\partial^2 u}{\partial t^2}\Big\rangle dt = \int_0^T\Big\langle m\frac{\partial^2\tilde{\lambda}}{\partial t^2}, u\Big\rangle dt - \Big\langle m\frac{\partial\tilde{\lambda(T)}}{\partial t}, u(T)\Big\rangle + \Big\langle m\frac{\partial\tilde{\lambda(0)}}{\partial t}, u(0)\Big\rangle + \tag{A.13}$$

$$\Big\langle\tilde{\lambda(T)}, m\frac{\partial u(T)}{\partial t}\Big\rangle - \Big\langle\tilde{\lambda(0)}, m\frac{\partial u(0)}{\partial t}\Big\rangle \tag{A.14}$$

Now computing the

$$\frac{\partial L(u,\tilde{\lambda},m)}{\partial u} = 0 \tag{A.15}$$

we get the following set of adjoint equations

$$\begin{cases}\tilde{\lambda}(T) = 0\\ \frac{\partial\tilde{\lambda}}{\partial t}(T) = 0\\ m\frac{\partial^2\tilde{\lambda}}{\partial t^2} - \nabla^2\tilde{\lambda} = P^\dagger(Pu - d)\end{cases} \tag{A.16}$$

The $\tilde{\lambda}$ is called as adjoint wavefield where the source term is the residual of the data recorded at receiver . The another thing to note is that the system has final boundary conditions at $t = T$. To give a physical meaning we introduce a change in variables $t = T - t'$

$$\begin{cases}\tilde{\lambda}(0) = 0\\ \frac{\partial\tilde{\lambda}}{\partial t'}(0) = 0\\ m\frac{\partial^2\tilde{\lambda}}{\partial t'^2} - \nabla^2\tilde{\lambda} = P^\dagger(Pu(T-t') - d(T-t')\end{cases} \tag{A.17}$$

Now the residuals are reverse in time and are back-propagated. The gradient now finally is

$$\frac{\partial J}{\partial m} = -\int_{t=0}^{t=T} \tilde{\lambda}(T-t)\frac{\partial^2 u}{\partial t^2} \tag{A.18}$$

In terms of velocity since $m = \frac{1}{v^2}$

$$\frac{\partial J}{\partial v} = \frac{2}{v^3}\int_{t=0}^{t=T} \tilde{\lambda}(T-t)\frac{\partial^2 u}{\partial t^2} \tag{A.19}$$

Appendix B

Reverse Time Migration

The imaging condition of Reverse time migration is same as Full Waveform Inversion . The only difference is the data is backpropagated instead of residuals.

$$I(x) = \sum_{s=1}^{s=N_s} \int_{t=0}^{t=T} \sum_{r=1}^{s=N_r} u_s(x,t) u_r(x,t) dt \tag{B.1}$$

where $u_s(x,t)$ is the source wavefield and $u_r(x,t)$ is the backpropagated receiver wavefield. Reverse Time migration seeks an image with high wavenumber content, and low wavenumbers represent artefacts in the image. Low fequency artefacts are created in the presence of strong reflectors by the correlation of wavefields with back-scattered energy. We also included the source illumination term in the imaging condition . The imaging condition with source illumination is as

$$I(x) = \sum_{s=1}^{s=N_s} \frac{\int_{t=0}^{t=T} \sum_{r=1}^{s=N_r} u_s(x,t) u_r(x,t) dt}{\int_{t=0}^{t=T} u_s(x,t) u_s(x,t) dt + \epsilon^2} \tag{B.2}$$

ϵ^2 is a small number introduced to prevent singularity.

We get rid of direct waves by introducing a water layer above in our model. The parallelisation strategies that we discussed for Full waveform inversion are equally applicable for Reverse Time Migration as well.

Bibliography

[1] A. Fichtner, *Full seismic waveform modelling and inversion*. Springer Science & Business Media, 2010.

[2] S. Operto, Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux, "A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice," *The Leading Edge*, vol. 32, no. 9, pp. 1040–1054, 2013.

[3] R. G. Pratt, C. Shin, and G. Hick, "Gauss–newton and full newton methods in frequency–space seismic waveform inversion," *Geophysical Journal International*, vol. 133, no. 2, pp. 341–362, 1998.

[4] M. J. Woodward, "Wave-equation tomography," *Geophysics*, vol. 57, no. 1, pp. 15–26, 1992.

[5] L. Sirgue and R. G. Pratt, "Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies," *Geophysics*, vol. 69, no. 1, pp. 231–248, 2004.

[6] L. Sirgue, R. G. Pratt *et al.*, "Waveform inversion under realistic conditions: Mitigation of non-linearity," in *2003 SEG Annual Meeting*. Society of Exploration Geophysicists, 2003.

[7] P. Mora, "Inversion= migration+ tomography," in *Parallel Computing 1988*. Springer, 1989, pp. 78–101.

[8] D. B. Kirk and W. H. Wen-mei, *Programming massively parallel processors: a hands-on approach*. Newnes, 2012.

[9] V. Khaidukov, E. Landa, and T. J. Moser, "Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution," *Geophysics*, vol. 69, no. 6, pp. 1478–1490, 2004.

[10] P. Sava and B. Biondi, "Wave-equation migration velocity analysis. i. theory," *Geophysical Prospecting*, vol. 52, no. 6, pp. 593–606, 2004.

[11] C. Chapman, *Fundamentals of seismic wave propagation*. Cambridge University Press, 2004.

[12] C. M. R. Fowler, *The solid earth: an introduction to global geophysics*. Cambridge University Press, 1990.

[13] V. Cerveny, *Seismic ray theory*. Cambridge university press, 2005.

[14] K. D. Klem-Musatov, F. Hron, L. R. Lines, and C. A. Meeder, *Theory of seismic diffractions*. Society of Exploration Geophysicists Tulsa, USA, 1994.

[15] W. S. Harlan, J. F. Claerbout, and F. Rocca, "Signal/noise separation and velocity estimation," *Geophysics*, vol. 49, no. 11, pp. 1869–1880, 1984.

[16] E. Landa, V. Shtivelman, and B. Gelchinsky, "A method for detection of diffracted waves on common-offset sections*," *Geophysical Prospecting*, vol. 35, no. 4, pp. 359–373, 1987.

[17] E. Landa and S. Keydar, "Seismic monitoring of diffraction images for detection of local heterogeneities," *Geophysics*, vol. 63, no. 3, pp. 1093–1100, 1998.

[18] S. Fomel, "Applications of plane-wave destruction filters," *Geophysics*, vol. 67, no. 6, pp. 1946–1960, 2002.

[19] ——, "Shaping regularization in geophysical-estimation problems," *Geophysics*, vol. 72, no. 2, pp. R29–R36, 2007.

[20] J. F. Claerbout, *Earth soundings analysis: Processing versus inversion*. Blackwell Scientific Publications Cambridge, Massachusetts, USA, 1992, vol. 6.

[21] M. T. Taner, S. Fomel *et al.*, "Separation and imaging of seismic diffractions using plane-wave decomposition," in *2006 SEG Annual Meeting*. Society of Exploration Geophysicists, 2006.

[22] E. Landa, S. Fomel, M. Reshef *et al.*, "Separation, imaging, and velocity analysis of seismic diffractions using migrated dip-angle gathers," in *2008 SEG Annual Meeting*. Society of Exploration Geophysicists, 2008.

[23] A. Klokov and S. Fomel, "Separation and imaging of seismic diffractions using migrated dip-angle gathers," *Geophysics*, vol. 77, no. 6, pp. S131–S143, 2012.

[24] E. Kozlov, N. Barasky, E. Korolev, A. Antonenko, E. Koshchuk *et al.*, "Imaging scattering objects masked by specular reflections," in *74th SEG Annual Meeting*, 2004, pp. 1131–1134.

[25] T. Moser and C. Howard, "Diffraction imaging in depth," *Geophysical Prospecting*, vol. 56, no. 5, pp. 627–641, 2008.

[26] B. Engquist and A. Majda, "Absorbing boundary conditions for numerical simulation of waves," *Proceedings of the National Academy of Sciences*, vol. 74, no. 5, pp. 1765–1766, 1977.

[27] C. Cerjan, D. Kosloff, R. Kosloff, and M. Reshef, "A nonreflecting boundary condition for discrete acoustic and elastic wave equations," *Geophysics*, vol. 50, no. 4, pp. 705–708, 1985.

[28] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *Journal of computational physics*, vol. 114, no. 2, pp. 185–200, 1994.

[29] J. M. Carcione, G. C. Herman, and A. Ten Kroode, "Seismic modeling," *Geophysics*, vol. 67, no. 4, pp. 1304–1325, 2002.

[30] J. A. Roden and S. D. Gedney, "Convolutional pml (cpml): An efficient fdtd implementation of the cfs-pml for arbitrary media," *Microwave and optical technology letters*, vol. 27, no. 5, pp. 334–338, 2000.

[31] G. Festa, E. Delavaud, and J.-P. Vilotte, "Interaction between surface waves and absorbing boundaries for wave propagation in geological basins: 2d numerical simulations," *Geophysical Research Letters*, vol. 32, no. 20, 2005.

[32] P. Williamson, "Tomographic inversion in reflection seismology," *Geophysical Journal International*, vol. 100, no. 2, pp. 255–274, 1990.

[33] G. T. Schuster, "Basics of seismic wave theory," *de Notes for the Lecture Courses, University of Utah*, 2007.

[34] R.-E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 2006.